

# Sussex Research Online

## Red, yellow, green and blue are not particularly colorful

Article (Accepted Version)

Witzel, Christoph, Maule, John and Franklin, Anna (2019) Red, yellow, green and blue are not particularly colorful. *Journal of Vision*. ISSN 1534-7362 (Accepted)

This version is available from Sussex Research Online: <http://sro.sussex.ac.uk/id/eprint/88862/>

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

### **Copyright and reuse:**

Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

# Red, yellow, green and blue are not particularly colorful

**Christoph Witzel**

Experimental Psychology, Justus-Liebig-  
University, Gießen, Germany



**John Maule**

School of Psychology, University of Sussex,  
Brighton, UK

**Anna Franklin**

School of Psychology, University of Sussex,  
Brighton, UK



Colorfulness and saturation have been neglected in research on color appearance and color naming. Perceptual particularities, such as cross-cultural stability, “focality”, “uniqueness”, “salience” and “prominence” have been observed for red, yellow, green, and blue, when those colors were more saturated than other colors in the stimulus samples. The present study tests whether high saturation is a particular property of red, yellow, green and blue, which would explain those observations. First, we carefully determined the category prototypes and unique hues for red, yellow, green, and blue. Using different approaches in two experiments, we assessed discriminable saturation as the number of just-noticeable differences away from the adaptation point (i.e. neutral gray). Results show that some hues can reach much higher levels of maximal saturation than others. However, typical and unique red, yellow, green, and blue are not particularly colorful. Many other, intermediate colors have a larger range of discriminable saturation than these colors. These findings suggest that prior claims of perceptual salience of category prototypes and unique hues actually reflect biases in stimulus sets rather than perceptual properties. Additional analyses show that consistent prototype choices across fundamentally different languages are strongly related to the variation of discriminable saturation in the stimulus sets. Our findings also undermine the idea that every color can be produced by a mixture of unique hues. Finally, the measurements in this study provide a large amount of data on saturation across hues, which allows for reevaluating existing estimates of saturation in future studies.

**Keywords:** color categorization, color discrimination, color naming, color vision, Sapir-Whorf Hypothesis, saturation

## Introduction

Colorfulness might provide the missing link between color perception, color appearance, and color naming, a central topic in color science and in research on the relationship between perception and language (for review, see Lindsey & Brown, 2019; Siuda-Krzywicka et al., 2019; Witzel, 2018a; Witzel & Gegenfurtner, 2018b). *Colorfulness* is the attribute of a perceived color according to which the color appears to be more or less chromatic. In other words, it refers to the difference of a color from achromatic colors, such as black, white, and gray. A more precise distinction may be made between colorfulness, chroma and saturation depending on whether colorfulness is assessed relative to the brightness of the adapting white-point (*chroma*) or relative to the brightness of the chromatic stimulus itself (*saturation*) (Fairchild, 2005, p. 87ff).

## Background

Colorfulness plays a major, yet widely neglected role in the investigation of color naming and color appearance (for review, see Witzel, 2018b). In color naming, the multitude of perceivable colors are grouped into color categories by a few basic color terms. For example, English color terms define three achromatic (*black*, *gray*, and *white*) and eight chromatic categories (*pink*, *red*, *orange*, *yellow*, *green*, *blue*, *purple*, and *brown*). Each category contains a *prototype*, i.e. the most typical color of the category, for example the red that is redder than any other red.

Color appearance refers to how colors subjectively appear to the observer. It has been suggested that the appearance of any color is a combination of *unique hues* (for review, see e.g. Abramov & Gordon, 1994; Valberg, 2001). Apart from black and white, these unique hues correspond to pure red, yellow, green, and blue, where “pure” means that the hue

does not contain any of the other hues. For example, unique red is neither yellowish nor bluish (nor whitish or blackish). The idea of unique hues follows Hering's (1964 [1878]) idea of *primary colors* (or *Urfarben*). According to this idea all colors are defined by two chromatic opponent color pairs (unique yellow and blue, and unique red and green), and one achromatic color pair (black and white).

A large range of studies suggests that unique hues and the prototypes of the English categories red, yellow, green and blue have particular perceptual properties. Several studies found statistical regularities in color naming and categorization across fundamentally different languages and cultures (Berlin & Kay, 1969; Kay & Regier, 2003; Lindsey & Brown, 2009, 2006; Lindsey, Brown, Brainard, & Apicella, 2015; Gibson et al., 2017). The prototypes of the English color categories are particularly stable across languages (Berlin & Kay, 1969; Regier, Kay, & Cook, 2005; Webster et al., 2002). It has also been shown that those prototypes are easier to name and memorize, even across cultures with different languages. To explain the particular role of English prototypes it has been suggested that the prototypes of the English color categories are particularly "salient" and "linguistically codable" (Rosch Heider, 1972; Brown & Lenneberg, 1954; Hays, Margolis, Naroll, & Perkins, 1972; Witkowski & Brown, 1982; Boynton & Olson, 1990; Sturges & Whitfield, 1997; Bolton, 1978). Under the assumption that English category prototypes have a particular property due to which they correspond to the "focus" of universal color categories independent of language they were termed *focal colors* (Rosch Heider, 1972; Berlin & Kay, 1969; for review see Witzel, 2018a,b).

Regier and colleagues (2007) showed that the different categories of a wide range of languages were distributed so that the color chips within the respective categories tended to be more similar than those across categories. The authors argued that the high similarity around the category centers showed that focal colors are perceptually salient (in a broad sense), and that categories developed around these perceptually salient colors. Another study found that unique hues are *perceptually prominent* (Kuehni, Shamey, Mathews, & Keene, 2010). Based on observers' judgments about color similarity, the study provided evidence that unique hues subjectively appear more different from one another than intermediate hues. This observation suggests that unique hues "stick out" in color appearance, and hence are perceptually prominent. Studies using a technique called *partial hue matching* found that intermediate hues subjectively appear to be similar to both adjacent unique hues, while unique hues appear to be completely different from one another (Logvinenko, 2012; Logvinenko & Beattie, 2011). Finally, several studies provided evidence that both category prototypes and unique hues correspond to surfaces that reflect light in a way that is particularly predictable (*singular*) across illuminations (Philipona & O'Regan, 2006; Vazquez-Corral, O'Regan, Vanrell, & Finlayson, 2012).

All the above findings suggest that focal colors may act as *perceptual anchors*, i.e. points of reference in color space that are stable across observers and illuminations and around which color appearance and color categorization are organized (Witzel, Maule, & Franklin, 2013; Witzel, Cinotti, & O'Regan, 2015). At the same time, all those studies used maximally saturated Munsell chips. In this set of Munsell chips, saturation strongly varies across hue and lightness because maximum Munsell Chroma is not constant across hue and lightness. The Munsell Color System provides particularly high degrees of Munsell Chroma at and close to typical red, yellow, green, and blue (Collier, 1973; Collier et al., 1976; Jameson & D'Andrade, 1997; Boynton, MacLaury, & Uchikawa, 1989). As a result, the set of maximally saturated Munsell chips tends to have local peaks of saturation at or close to typical and unique red, yellow, green, and blue (cf. Figure 4.a-b in Witzel et al., 2015; Figure 1 in Witzel, 2018b). The peaks of saturation around typical chips provide an alternative explanation for the findings in support of their perceptual salience:

First of all, saturation and lightness determine visual salience when the background is gray because visual salience is defined as the contrast of a color to its background (Itti, 2007). If the purest, most typical red, yellow, green, and blue were particularly colorful they would "jump out to the eye" (cf. Figure 1 in Witzel, 2018). Visual salience might be the reason why observers strongly respond to those colors independent of culture and language.

Second, color categorization depends on saturation (Witzel, 2018b). Color naming is less consistent for desaturated than for highly saturated or achromatic colors, and observers almost never choose a desaturated color as a category prototype (e.g. Figure 8 in Olkkonen, Witzel, Hansen, & Gegenfurtner, 2010). As a result, there is a correlation between category consistency and Munsell Chroma (Figure 2 in Witzel, 2018b). Observers tend to choose more saturated colors as category prototypes and unique hues (Witzel, 2019). The variation of saturation in the set of maximally saturated Munsell chips is correlated with the cross-cultural prototype choices (Figure 5.c in Witzel et al., 2015), and with categorization in non-industrialized, remote cultures (Lindsey et al., 2016; Witzel, 2016). Hence, the peaks of saturation around typical and unique red, yellow, green, and blue in this stimulus set may explain the observed regularities in categorization across different languages.

Third, if prototypes and unique hues are more saturated than intermediate hues, they are further away from the white-point. As a logical necessity, distances between points increase when increasing their distance to the origin (cf. Witzel, 2018, p.50). For this reason, the saturated colors identified as prototypes and unique hues must have higher distances to one another than intermediate hues (Logvinenko, 2012; Logvinenko & Beattie, 2011; Kay & Regier, 2007; Kuehni et al., 2010).

Finally, the sensory singularity of prototypes and unique hues completely disappears when using more uniformly saturated instead of maximally saturated Munsell chips (Witzel et al., 2015).

Taken together, peaks in saturation around typical red, yellow, green, and blue are likely to be the origin of the particular characteristics found for category prototypes and unique hues. However, these peaks were due to sampling stimuli from the set of maximally saturated Munsell chips. The limits of maximal saturation in the Munsell system are at least partially due to pigments, and do not necessarily reflect properties of the perceptual system (Pastilha, Linhares, Rodrigues, & Nascimento; cf. Figure 4 Witzel, 2018b). Hence, it is not clear whether focal colors are actually more saturated than any other colors. The particularities found for typical and unique red, yellow, green, and blue might well be an artifact of the particular stimulus choice.

## Objective

The present study investigates whether the hues that correspond to typical and unique red, yellow, green, and blue have higher levels of colorfulness than intermediate colors. For this purpose, we compared the maximum colorfulness across hues at the *visible gamut*. The visible gamut is the limit of (visible) chromaticities and corresponds to the chromaticities of spectral colors. For the comparison across hues we fixed lightness at the lightness typical for red, yellow, green, and blue, respectively. Note that the differences between chroma and saturation are unimportant for this study because comparisons were done at equal lightness for each hue range.

However, colorfulness is difficult to measure since there are no metrics that allow for reliable quantification. Chroma (and saturation) can be roughly approximated by the radius (and radius divided by lightness) in color appearance models, such as CIELUV, CIELAB, CIECAM02, and in the Munsell system (e.g. Fairchild, 2005, pp. 189-190). Each of these models gives quite different predictions of chroma across hues, and comparisons across hues and lightness may be strongly affected by the inhomogeneity of these color spaces.

To assess maximum perceivable saturation, we determined how many different levels of saturation an observer is able to perceive between achromatic gray and the saturation at the visible gamut. Following the idea of Fechnerian discrimination scaling (e.g. Dzhafarov & Colonius, 2011; Irtel, 2014), we quantified “levels of saturation” through discrimination thresholds for differences in saturation. For this, we measured just-noticeable differences of CIELUV saturation, which is a function of colorimetric purity. A just-noticeable difference, or JND, is the minimum difference between two levels of saturation that an observer is just able to see. To estimate the saturation of a given color, we counted how many JNDs fit between that color and the

neutral chromaticity to which the observer is adapted (i.e. gray). We call this measure of saturation *discriminable saturation* because it is based on discriminability (e.g. Dzhafarov & Colonius, 2011; Irtel, 2014). Note that the number of JNDs is independent of the color space, in which JNDs are measured. We counted the number of JNDs between adapting gray and the visible gamut to estimate the maximum perceivable saturation in terms of discriminable saturation. We call this maximum discriminable saturation *visible saturation*.

The just-noticeable differences also allow us to assess the *sensitivity to saturation*. It has been shown that JNDs are linearly related to the saturation (radius) of the cone-opponent channels in DKL-space (Krauskopf & Gegenfurtner, 1992). A linear relationship allows for calculating Weber fractions as a measure of sensitivity that is independent of axis scaling. We tested whether the sensitivity to saturation and the visible saturation was higher for category prototypes and unique hues than for non-typical and intermediate colors.

The investigation of discriminable saturation and of the sensitivity to saturation complements our investigations of *subjective saturation* in a companion study (Witzel & Franklin, 2014). While discriminable saturation assesses how many steps of saturation can be discriminated, subjective saturation describes how saturated a color subjectively appears to an observer. Witzel & Franklin (2014) had measured subjective saturation through saturation matches across hue, and tested whether subjective saturation was related to unique hues and category membership. The extensive data on discriminable saturation from the present study also allows for re-evaluating how subjective saturation relates to discriminable, i.e. whether colors of equal subjective saturation imply equal levels of discriminable saturation, or whether these two measures of saturation dissociate.

We first determined typical lightness and hue in preliminary measurements, which were needed for the subsequent measurements of saturation. *Typical lightness* and *typical hue* are the lightness and hue of a category prototype. Second, we measured JNDs of saturation differences at different levels of saturation across the monitor gamut (Experiment 1). These measurements involved a high range of saturation levels across the monitor gamut and allowed for extrapolating JNDs towards the visible gamut. We then compared sensitivity and discriminable saturation between typical and non-typical hues. These measurements complete the preliminary results of a conference paper (Witzel, Maule, & Franklin, 2013). Finally, we measured sensitivity and visible saturation for a larger range of hues in order to inspect the local variation of saturation across hues after accounting for global trends across hue (Experiment 2).



## Lightness and hue

These preliminary measurements were aimed at determining the typical lightness and the typical and unique hue for red, yellow, green and blue. The perception of a hue may change with saturation due to the Abney effect (Burns, Elsner, Pokorny, & Smith, 1984; O'Neil et al., 2012; Mizokami, Werner, Crognale, & Webster, 2006). For this reason, we examined whether unique hues systematically change across different levels of saturation. We also tested whether there is a difference between unique hues and the typical hues of red, yellow, green, and blue.

### Method

We first measured the typical lightness of each category prototype. Using the results of those measurements, we then determined typical and unique hues at the respective lightness levels. We measured typical hues along an isoluminant hue circle in CIELUV, and unique hues along three hue circles with different CIELUV chroma. To reach high levels of saturation within the monitor gamut, we also determined unique hues in sections of highly saturated hue circles that were within the gamut in the region around the targeted unique hues, but crossed the gamut for other unique hues. We call these latter “super saturated” measurements. We established category boundaries through color naming. **Tables 1-2** provide overviews of the different kinds of typical and unique hue measurements.

### Participants

Six observers (3 women; 4 British, 1 German, 1 Chinese; mean age 24.4y, SD = 6.1y) took part in the measurements of typical lightness. Overall 44 observers (30 women, age:  $25.4 \pm 6.4$ y) participated in the measurements of category boundaries, typical and unique hues. Details on sample sizes of each condition may be found in **Table 2** (column “N”). In addition, N = 5 observers adjusted prototypes and unique hues at “extra dark” lightness. All observers except the first author (cw) were naïve as to the purpose of the experiment, and were paid for participation. None of the observers was red-green color deficient, as verified through Ishihara plates (Ishihara, 2004). Ethics approval was granted by the Life Sciences and Psychology Cluster based Ethics Committee at the University of Sussex.

### Apparatus

Stimuli were displayed on a Mitsubishi Diamond Pro 2070SB CRT monitor driven by a NVIDIA graphics card (NVIDIA Corporation, Santa Clara, CA) with a color resolution of 8 bits per channel, and a spatial resolution of  $1600 \times 1200$  pixels. The refresh-rate was 80Hz and the CIE1931 chromaticity coordinates and luminance of the monitor primaries were  $R = (0.626, 0.337, 13.01)$ ,  $G = (0.283, 0.612, 47.8)$ , and  $B = (0.151, 0.071, 8.0)$ . Gamma corrections without bit loss were applied based on the measured gamma curves of the monitor primaries. Observers

used a chin rest to control a viewing distance of 80 cm to the screen. Experiments were conducted in a black room (black-painted, window-less, all potential light sources covered), and observers looked through a black viewing tunnel in order to guarantee that they adapted to the background of the computer screen only. Initial adaptation was accomplished by presenting instructions and initial practice trials on the gray screen in all tasks. A separate number pad was used to register responses. The apparatus was the same in all experiments of this study.

### Stimuli

In all tasks colors were presented as colored disks of 3.2 deg visual angle. At each lightness level, test colors were sampled along an isoluminant hue circle in CIELUV space (cf. **Figure S1**). A CIELUV hue circle has the advantage that it roughly controls perceived color distances (cf. Figure 1.a in Witzel & Gegenfurtner, 2018a; for review see Fairchild, 2005).

The adapting white-point was defined as standard illuminant C ( $x = 0.3101$ ,  $y = 0.3162$ ) of  $50\text{cd/m}^2$ . The gray background had the chromaticities of the white-point, but was set to  $L^* = 70$  ( $Y = 20.4\text{cd/m}^2$ ) to allow for displaying colors lighter than the background. The chromaticities of standard illuminant E ( $x = 0.3333$ ,  $y = 0.3333$ ) were used for measurements of dark colors at  $L^* = 50$ . This was done to better exploit the monitor gamut at this lightness and made it possible to sample hues at a radius of 50 at  $L^* = 50$  (instead of only 45 with illuminant C). **Table 1** provides an overview of these settings.

**Lightness measurements.** Saturation necessarily changes across lightness, converging towards zero when approaching black and white (i.e. maximum lightness). For this reason, saturation could not be completely controlled across lightness levels. We fixed CIELUV chroma at a radius of 50 for lightness levels at which such a radius existed within the monitor gamut that is for  $L^* = [56, 82]$ . For all other lightness levels, the CIELUV radius was defined by a circle that was tangential with the monitor gamut. Hence at low and high lightness, a change of lightness implied a change of saturation. We limited lightness adjustments to a minimum of  $L^* = 20$ , and a maximum of  $L^* = 94$ , at which the maximum radius was 18 and 15, respectively. This was done to make sure that participants could identify hues at all adjustable lightness levels.

**Category prototype and boundary measurements.** CIELUV radius was determined so that the circle was as large as possible without transgressing the monitor gamut. For details see column “Max chroma” of **Table 1**. The stimulus sets for the color naming measurements at each lightness level consisted of 120 test colors uniformly sampled along the isoluminant hue circle (from 0 to 357 in 3 deg steps).

**Main unique hue measurements.** Hues were sampled at three different levels of saturation, namely at CIELUV radii of 30, 40, and 50 for  $L^* = 50$  and  $L^* = 76$ ; of 35, 45, and 55

for  $L^* = 60$ ; and of 14, 24 and 34 for  $L^* = 38$ . These measurements involved all hues along the three isoluminant hue circles.

**Super saturated unique hue measurements.** The range of hues that could be adjusted was limited to an area with only the most relevant hues, e.g. only reddish hues when adjusting red. This made it possible to achieve higher CIELUV radii by exploiting protrusions of the monitor gamut. These measurements were not conducted for each lightness level. Instead, colors were always presented at the typical lightness of the target descriptor ( $L^* = 76$  for yellow,  $L^* = 60$  for blue, and  $L^* = 50$  for red and green).

**Table 1.** Color specifications. Purpose = the aim of the measurements;  $L^*$  = CIELUV lightness; Y = luminance in Candela per square meter; Max chroma = Maximum CIELUV chroma achievable within monitor gamut; WP = white-point; C = illuminant C with  $xyY = [0.3101, 0.3162, 50]$ ; E = illuminant E with  $xyY = [0.3333, 0.3333, 50]$ . Background lightness was always  $L^* = 70$  ( $20.4 \text{ cd/m}^2$ ).

| Purpose     | $L^*$ | Y<br>[cd/m <sup>2</sup> ] | Max<br>chroma | WP |
|-------------|-------|---------------------------|---------------|----|
| Lightness   | 20-94 | 1.5-42.6                  | 15-50         | C  |
| Yellow      | 76    | 24.9                      | 50            | C  |
| Blue        | 60    | 14.1                      | 55            | C  |
| Red & Green | 50    | 9.2                       | 50            | E  |
| Extra dark  | 38    | 5.0                       | 34            | C  |

### Procedure

**Lightness measurements.** Participants were asked to adjust the lightness and hue of the prototypes of all 8 chromatic basic color terms (pink, red, orange, yellow, green, blue, purple, and brown). Each prototype was adjusted 5 times. Each key press for lightness adjustments produced changes in lightness of  $L^* = 2$ . When observers reached the minimum and maximum lightness, a message appeared, indicating that they could not go further.

**Hue adjustment tasks (at fixed lightness).** Observers were asked to adjust the hue of a disk so that it corresponds to the respective target descriptions (i.e. a category prototype or unique hue). At the beginning of each trial the target description was presented. Then, a randomly colored disk was shown in the center of the screen. Two keys (keys 1 and 2) allowed observers to change hue in one and the other azimuth direction along the hue circle. Each single key press changed hue by 1 deg. However, observers could smoothly surf through the hue circle by holding another key (enter

key) while pressing the hue adjustment keys (keys 1 and 2). Once they reached the approximate region of the target description, they could fine-adjust hues through single key presses of the hue adjustment keys. If necessary, observers could also recall the target description through key press (key 0). At the end of each trial, observers confirmed their adjustment by key press (Back Space key).

For the **prototype adjustments**, participants were asked to adjust the color to match the most typical colour, or “best example”, of each of the eight chromatic basic color terms (red, yellow, green, blue, orange, pink, purple, and brown). There were 10 blocks. Each prototype was the target once in random order in each block. Observers could skip trials, in which no prototype could be found at the respective lightness and saturation of the measurement. However, the experimenter highlighted that this should only be done if they could not find any hue that fits the respective color description. For example, they should only skip pink if there was not even one example of pink among the test colors.

In the **unique hue** adjustments observers were instructed to adjust the hue that is neither one nor the other adjacent unique hue, for example the “yellow that is neither red nor green”. In these measurements, observers were also asked to adjust *binary hues* – hues that are 50% one and 50% the adjacent hue, for example “50% green and 50% blue”. These overall 8 target descriptions (4 unique and 4 binary hues) were done once in random order in each of 3 blocks. Unlike for the prototype adjustments, participants could not skip any trial of the unique hue adjustments.

In the **super saturated** unique hue adjustments, observers adjusted unique hues, only. Since small hue intervals constrain hue adjustments, red, green, and blue were measured at two saturation levels, where the lower saturation level allowed for larger hue intervals. Yellow was measured twice with the same radius and interval. Table 2 provides details on radii and intervals (see column “azi limits”). Each condition (4 unique hues and 2 radii) was done 5 times in each block, resulting in overall 40 adjustments per block.

In the **color naming task**, observers had eight keys available that corresponded to the eight chromatic basic color terms. After the presentation of a fixation point for 1 second, the colored disk was shown at the center of the screen until one of the 8 keys was pressed. In general, there were 10 blocks, in which each of the 120 colors was presented once, in random order. However, some observers did only 5 blocks (2 at  $L^* = 50$ , 4 at  $L^* = 60$ , and 9 at  $L^* = 76$ ), one observer did 15 blocks at  $L^* = 76$ , and observers cw and f1 did all these sessions twice (20 blocks for each lightness level).

**Table 2.** Typical and unique hue measurements. Typical = adjustments of prototypes; unique = adjustments of unique hues; super = adjustments of unique hues at high saturation. “Rad” radius (chroma), “azi” azimuth (hue) in degree; “azi limits” limits of interval within which azimuth could be adjusted; “–” no limits, adjustments along 360 deg; “adj azi” adjusted azimuth, “N” number of participants. Measurements with \* are pooled in Figure 3. “Combined” pools the data of all three kinds of measurements.

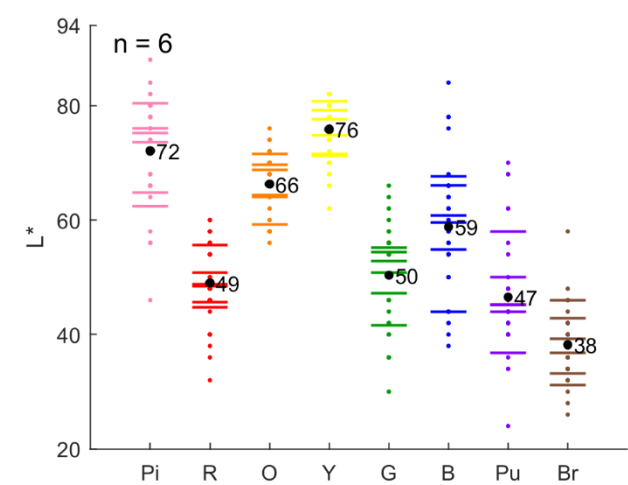
| RED ( $L^* = 50$ ) | YELLOW ( $L^* = 76$ ) | GREEN ( $L^* = 50$ ) | BLUE ( $L^* = 60$ ) |
|--------------------|-----------------------|----------------------|---------------------|
|--------------------|-----------------------|----------------------|---------------------|

| Task     | rad | azi limits | adj azi | N  | rad | azi limits | adj azi | N  | rad | azi limits | adj azi | N  | rad | azi limits | adj azi | N  |
|----------|-----|------------|---------|----|-----|------------|---------|----|-----|------------|---------|----|-----|------------|---------|----|
| Naming   | 50  | -          | -       | 17 | 50  | -          | -       | 17 | 50  | -          | -       | 17 | 55  | -          | -       | 10 |
| Typical  | 50  | -          | 17.5    | 10 | 50  | -          | 77.8    | 20 | 50  | -          | 130.6   | 12 | 55  | -          | 227.3   | 8  |
| Unique   | 30  | -          | 9.4     |    | 30  | -          | 79.4    |    | 30  | -          | 136.3   |    | 35  | -          | 219.2   |    |
|          | 40  | -          | 6.0     | 19 | 40  | -          | 79.6    | 18 | 40  | -          | 136.8   | 19 | 45  | -          | 228.4   | 17 |
|          | 50  | -          | 9.2     |    | 50  | -          | 83.1    |    | 50* | -          | 138.2   |    | 55* | -          | 225.5   |    |
| Super    | 67  | 319-41     | 9.6     | 16 | 89* | 49-107     | 78.2    | 16 | 50* | 67-213     | 133.8   | 16 | 55* | 169-263    | 221.2   | 16 |
|          | 97  | 346-24     | 11.8    |    | 89* | 49-107     | 77.2    |    | 55  | 110-170    | 136.9   |    | 64  | 220-254    | 227.6   |    |
| Combined |     |            | 9.5     | 24 |     |            | 78.6    | 33 |     |            | 135.4   | 24 |     |            | 225.0   | 20 |

Results and Discussion

Lightness

Figure 1 illustrates the results for all eight prototypes. The most important prototypes for the subsequent measurements were red, yellow, green, and blue. We aimed at a fixed lightness for all observers to allow for comparisons across observers and hues. For this reason, we determined the lightness by the average lightness adjustments across observers (black dots). To simplify, we rounded the  $L^*$  of red ( $L^* = 49$ ) and blue ( $L^* = 59$ ) to the next multiple of 10. This also allowed for using the same lightness level for red and green ( $L^* = 50$ ). The lightness of brown ( $L^* = 38$ ) was used as an additional “extra dark” lightness level to further clarify the dependence of hue measurements on lightness. These results defined the lightness levels for the hue measurements and correspond to the respective lightness specifications in Table 1.



**Figure 1.** Typical lightness adjustments. The different categories are listed along the x-axis; labels refer to English color terms (Pi = pink, R = red etc.). The y-axis represents lightness, measured as  $L^*$  in CIELUV. Dots correspond to single measurements, horizontal lines to the averages of each participant, and black dots to the overall average across participants. Black digits indicate the average  $L^*$ .

Typical and unique hues

Figure S2 in the [Supplementary Material](#) provides detailed individual data for all unique hues and prototypes and at all lightness levels. Figure 2 concentrates on aggregated results at the typical lightness levels of red (panel a), green (b), yellow (c), and blue (d).

**The role of saturation.** The white disks in Figure 2 refer to the unique hue adjustments at different levels of CIELUV chroma. We calculated correlations between radius and azimuth for each participant, applied a Fisher-transformation to the correlation coefficients, and tested with a two-tailed t-test across participants whether they were significantly different from zero. Reported average correlation coefficients are averaged as Fisher transforms and converted back to correlation coefficients. Detailed results of these tests are given in Table S1 of the [Supplementary Material](#). For red, yellow, and green (Figure 3.a-c) there was no significant difference from zero (min.  $p = 0.46$ ), indicating that there was no systematic change in hue adjustments with chroma. The average correlation coefficient for blue was  $r = 0.40$ , which was close to significance ( $t(17)=2.0$ ,  $p = 0.06$ ). A positive correlation would indicate that adjustments of unique blue shift towards higher azimuth (red) with higher chroma. Such a shift contradicts the Abney effect (Burns et al., 1984; O’Neil et al., 2012; Mizokami et al., 2006). Moreover, the observed shift seems to be a particularity of the lowest level of saturation (35) rather than an overall trend of hue adjustment for increasing saturation (cf. Figure 2.d and Table 2). Instead of the Abney effect, a shift of blue stimuli could also be due to the fact that CIELUV space uses CIE1931 color matching functions (CIE, 1932), which underestimate sensitivity in the short-wavelength part of the spectrum compared to the more precise Judd (Judd, 1951) or Judd-Voss (Vos, 1978) corrected color matching functions.

**Typical vs. unique hues.** The vertical lines in Figure 2 indicate category boundaries (solid, black) and prototypes (colored, dotted). We averaged unique hue adjustments across saturation levels, and tested with independent t-tests across participants whether unique hues (e.g. the red that is neither blue nor yellow) were significantly different from typical hue adjustments (e.g. the best example of red). Table S2 in the [Supplementary Material](#) provides detailed results.

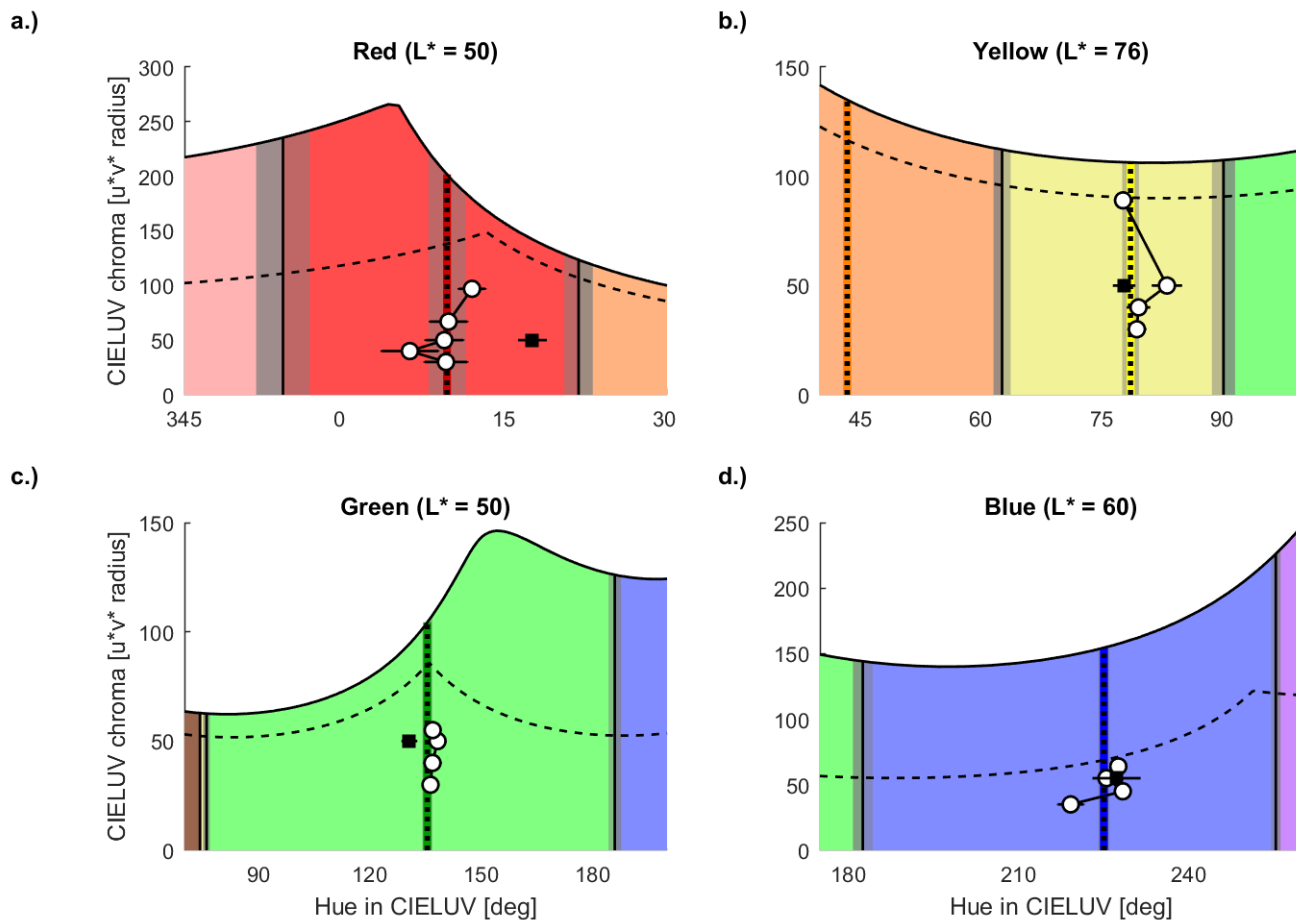
Typical and unique yellow (77.8 vs. 79.7 deg), and blue (227.8 vs. 224.8 deg) barely differed, and their differences were not statistically significant (both  $p > 0.38$ ). However, typical red settings were significantly closer to orange than unique red settings (17.5 vs. 7.9 deg,  $t(29) = 2.5$ ,  $p = 0.02$ , and typical green settings were closer to yellow than unique green settings (130.6 vs. 136.9 deg,  $t(31) = -2.1$ ,  $p = 0.049$ ).

**Variability of red.** Measurements of unique red, typical red and the red category boundaries seemed to vary much more strongly than the respective measurements in the other categories (see error bars across observers in **Figure 2.a** compared to **Figure 2.b-d**). A particularly high variability of red has also been observed in previous measurements along an isoluminant hue circle and was attributed to high lightness and low saturation in the stimulus samples (Witzel & Gegenfurtner, 2013; Hansen, Walter, & Gegenfurtner, 2007). Examinations of the color naming and skipped trials in typical adjustments revealed that observers did not reliably identify red colors, even at its typical lightness  $L^* = 50$ . Red category boundaries in color naming strongly varied because observers identified different categories (purple, pink, red, orange, and brown) in the reddish hue region (cf. **Figure S2.c**). In particular, the boundaries of red depended on whether observers found pink and orange colors, or whether they located red directly adjacent to purple and brown. Three of 17 observers did not even identify any red color.

**Table S3** provides details on the number of skipped trials. Skipped trials in the prototype adjustments indicate that observers did not find any hue that matches the respective color category at a given level of lightness and saturation. Yellow, green, and blue were barely skipped at their typical lightness (white entries in **Table S3**). As would be expected from the distribution of color categories across lightness (cf. Figure 4 in Witzel & Gegenfurtner, 2018b), skipped rates for yellow increased for lower lightness, while green and blue were barely skipped at any lightness levels. In contrast, red was skipped in 25% of trials at the typical lightness level ( $L^* = 50$ ). The rate of skipped trials for red increased for higher ( $L^* = 60-76$ ) and lower ( $L^* = 38$ ) lightness levels, suggesting that the comparatively high rate of skipped trials for red was not due to lightness, but to low saturation (cf. Witzel & Gegenfurtner, 2013). Hence, the reason for the variability of red measurements is likely the limited saturation of the hue circle, which in turn is due to the shape of the monitor (and visual) gamut.

**Summary.** For the main experiments that follow below, these observations indicate that typical and unique hues are similar one to another and rather stable across saturation levels. We will come back in the Discussions of the main experiments to the potential variation of blue across saturation, and the potential difference between typical and unique hues in the case of red and green.





**Figure 2.** Typical, unique, and boundary hues in CIELUV. Color categories (colored areas), typical (black squares) and unique hues (white disks) for red (panel a), yellow (b), green (c), and blue (d) are shown in polar coordinates of CIELUV space at typical lightness ( $L^*$  in title of panels). The x-axis represents hue (azimuth in deg), the y-axis CIELUV chroma ( $u^*v^*$  radius). Vertical black lines show category boundaries and dashed vertical lines the average across typical and unique hues (cf. “combined” in Table 2). The white disks connected by black curves show the different levels of CIELUV chroma for each unique hue measurement. The solid black curve above the colored areas indicates the visible gamut, and the dashed black curve the monitor gamut. Gray shadows around vertical lines as well as horizontal error bars around symbols represent standard errors of mean across individuals. *Note the systematic differences between typical and unique red (panel a), but not for any other hues.*

## Global peaks of saturation (Exp1)

In this first main experiment we tested whether sensitivity and visible saturation are larger for category prototypes than for colors at the category boundary. For this purpose, we determined JNDs of all saturation levels within the visible gamut in the respective hue direction. We measured JNDs almost exhaustively across the saturation levels in the monitor gamut. We then fitted a function to the *Threshold-versus-intensity* (TVI) data, where intensity is the difference of the test color from the adapting white point. Based on this function, we determined all JNDs in the visible gamut. Then, we counted the number of JNDs that fit between the adapting gray-point and the visible gamut to determine visible saturation. JNDs were measured for each of the four categories, red, yellow, green, and blue. Based on the above

measurements of typical lightness and hue, we determined for each observer their individual prototypes and the colors at the category boundary (*boundary colors*) in each of the two hue directions, resulting in 3 hues in each of the four stimulus sets.

## Method

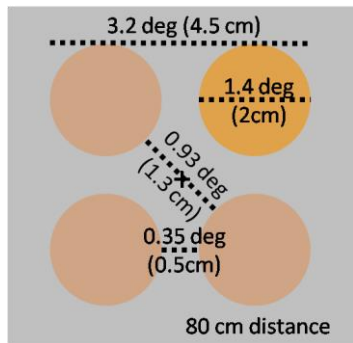
### Participants

Six observers participated in the measurements for the red and green categories (4 women,  $25.7 \pm 6y$ ), 8 observers (6 women, age:  $31.8 \pm 7y$ ) for yellow and 7 observers (5 women,  $25.4 \pm 5y$ ) for blue. An additional male observer (m0) was measured for yellow, but excluded from the main analyses due to color deficiencies. All participants were British apart from 1 German and 2 Chinese for the yellow, and 2 Germans and 1 Chinese for the other stimulus sets.

Observers in this experiment had participated in the measurements of typical and unique hues reported above.

### Stimuli

**Figure 3** illustrates the stimulus display in the discrimination task used to measure JNDs. It consisted of four colored discs, three of them showing the *test*, and one the *comparison color*. Disks had a size of 1.4 deg visual angle (2cm). Distances between disks were 0.35 deg visual angle (0.5cm) so that colors can be compared foveally, i.e. within 2 deg (max. 0.93 deg). The lowest spatial frequency of the stimuli (0.5 cpd) is below 1 cpd, which guarantees that contrast sensitivity is at ceiling and that the impact of spatial frequency on discrimination is minimal (Witzel & Gegenfurtner, 2015).

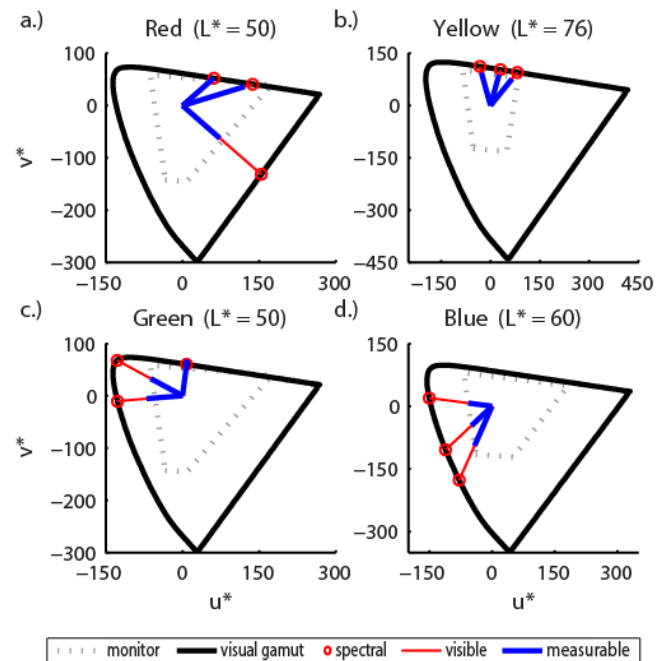


**Figure 3.** Stimulus display of 4AFC discrimination task. The 3 test colors are shown at lower saturation, the comparison color at higher saturation, as in descending staircases. The example illustrates orange-yellow hues from observer cw; but the colors in the figure may differ from those on the calibrated set-up. Distances are provided in visual angle (deg) and centimeters (cm). The 80cm distance is the distance of the observer from the screen. *Note that the maximum distance between two disks was still within the fovea (< 1deg).*

**Figure 4** illustrates the sampling of stimulus colors. As in the measurements of unique and typical hues, lightness ( $L^*$ ) was constant within a stimulus set. Test and comparison colors varied in saturation (CIELUV chroma) along the constant-hue lines away from the adapting gray-point (red-blue lines in **Figure 4**). The hue directions (azimuth) of the constant-hue lines were determined by the preliminarily measured typical and boundary hues (see first section “Lightness and hue”). We sampled the saturation of test colors so that adjacent test colors were about, or less than, one JND apart from each other. In this way, our measurement covered all discriminable levels of saturation within the monitor gamut (blue lines in **Figure 4**). Because the monitor gamut (gray line in **Figure 4**) is smaller than the visible gamut (thin black curve in **Figure 4**), JNDs between the monitor and the visible gamut were extrapolated (red line in **Figure 4**).

CW and f1 provided “exhaustive measures” of JNDs for all four stimulus sets, i.e. measures across several sessions for test colors at all saturation levels. Such exhaustive measures were also done with 3 more participants (f2-4) for the yellow

stimulus set only. All other datasets included at least 3 test colors, one at radius = 0 (detection threshold), one close to zero, and one at maximum available saturation within monitor gamut.



**Figure 4.** Stimulus sampling throughout the visible gamut in Experiment 1. Abscissa and ordinate correspond to  $u^*$  and  $v^*$ . Panels a, b, c, and d refer to the red, yellow, green, and blue categories at the typical lightness of the respective prototypes. The typical lightness ( $L^*$ ) is given in the title of the graphics. The red-blue lines indicate the hues of prototype (center line) and boundaries. The lines join at the origin, which corresponds to the neutral gray background. The black curve shows the visible gamut, the dotted gray one the monitor gamut. The blue part of the colored lines corresponds to the radius (chroma) of that hue that could be displayed on the monitor (i.e. were within the gray curve). The red part of those lines show the radius that could not be measured and had to be extrapolated in order to estimate the complete line in terms of JNDs. *Note that apart from green the prototypes do not coincide with the protrusions of the visible gamut CIELUV space (cf. Figure 2).*

### Procedure

JNDs were measured with a 4 Alternative-Forced Choice paradigm combined with a 3-up-1-down staircase. In this method, the comparison color (cf. **Figure 3**) is presented in a random location, and observers have to indicate which of 4 colored disks is different by pressing 1 of 4 keys that correspond to the locations of the four disks (for details see Krauskopf & Gegenfurtner, 1992; Witzel & Gegenfurtner, 2013). The 3-up-1-down staircase procedure decreases the difference between test and comparison, when responses were 3 times in a row correct, and increases the differences when a response was incorrect (for an illustration see Figure A1 in Witzel & Gegenfurtner, 2016; see also **Figure S3** in the [Supplementary Material](#)). This staircase converges to a

difference in saturation which corresponds to a probability of 0.79 of giving a correct response (Levitt, 1971), and a probability of 0.72 that the observer perceives this difference.

In order to measure JNDs in saturation, the test and comparison color were equal in lightness ( $L^*$ ) and hue (azimuth), and differed only in CIELUV chroma ( $u^*v^*$ ). For each test color, JNDs were measured for comparison colors with lower (increasing staircases) and higher (decreasing staircases) CIELUV chroma. These measurements were done in a separate block for each test color, with increasing and decreasing staircases interleaved. Each staircase stopped after five reversal-points. **Figure S3** illustrates example staircases.

The closer the difference between test and comparison is at the beginning of staircase, the faster the staircase converges and the more precise are the measurements. In order to optimize the staircase procedure, the starting colors of the staircases were determined through a color matching procedure. For this purpose, the stimulus display with the four discs was shown, and participants could adjust the saturation of the comparison. In a first block, they were asked to adjust it so that it is just visible. The location of the comparison color changed randomly with adjustment. To confirm the adjustment, they had to press the key that corresponds to the location of the odd one. In a second block, they were asked to adjust the saturation so that the disk was just invisible. The adjusted differences between test and comparison colors give a coarse estimation of the color differences above and below JNDs. These have been used to initialize staircases for the precise JND measurements that start above and below the estimated JNDs.

Measurements were carried out across several experimental sessions. In one session, JNDs were measured for 3 hues (1 typical and 2 boundary hues) at 3 levels of saturation, resulting in 9 blocks in random order. Each block included overall 4 interleaved staircases: decreasing and increasing staircases one of which started below and the other above the estimated JND (cf. **Figure S3**). Each experimental session took about 45-60min.

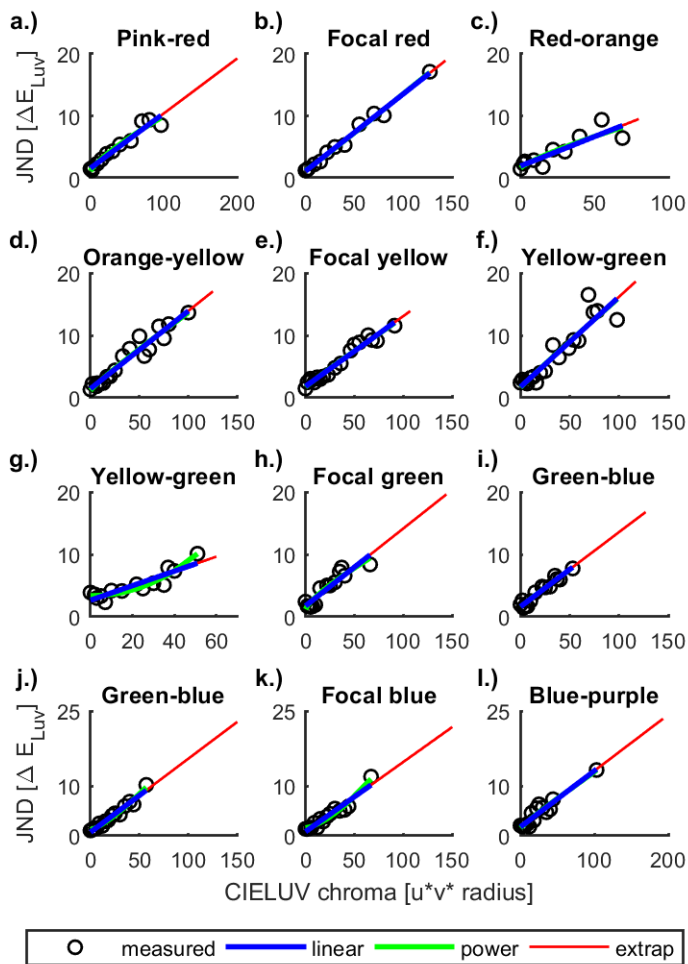
## Results

### JNDs

To determine JNDs, we discarded the first reversal-point to avoid artifacts and calculated the average over the remaining 4 (2 up and 2 down) reversal-points. **Figure 5** shows JNDs as a function of CIELUV chroma (radius in the  $u^*v^*$  plane) for participant f1. For all colors, JNDs clearly followed a linear distribution (blue line in **Figure 5**). In fact, a third-order polynomial fit resulted in almost the same fit as a simple line, as shown by the fact that the blue line almost completely covered the light green line in **Figure 5** (for other observers see **Figures S4-S5** in the [Supplementary Material](#); the interested reader may also note the linear trend for the additional, color-deficient observer m0 in **Figure S5.j-l**).

The clear linear trend of JNDs shows that these JNDs follow the Weber-Fechner Law in a broader sense (Since CIELUV is not a sensory color space, this trend cannot be considered as the Weber-Fechner law in a narrow sense). The linear trend allowed for calculating a Weber fraction as the slope of the linear function, following Weber's formula. This Weber fraction assesses sensitivity independent of the scaling of intensity, i.e. independent of the CIELUV radius. This makes it possible to calculate a metric of perceived intensity (or *JND-space*), in which 1 unit corresponds to 1 JND, following Fechner's discrimination scaling (for review see e.g. Dzhaferov & Colonius, 2011; Irtel, 2014). Based on this JND-space, we determined discriminable saturation at all levels of CIELUV radius, including the visible saturation at the visible gamut.

To do so, we fitted a linear function to the JNDs, using a least squares method. The Weber Fraction (the slope of the line) and the detection threshold (the intercept of the line) are a measure of sensitivity, and make it possible to compare sensitivity to differences in saturation across hues. The linear function also allowed us to extrapolate JNDs to the visible gamut (red lines in **Figure 5**) in order to calculate the visible saturation of each hue. These computations were done for the JNDs of each observer separately.



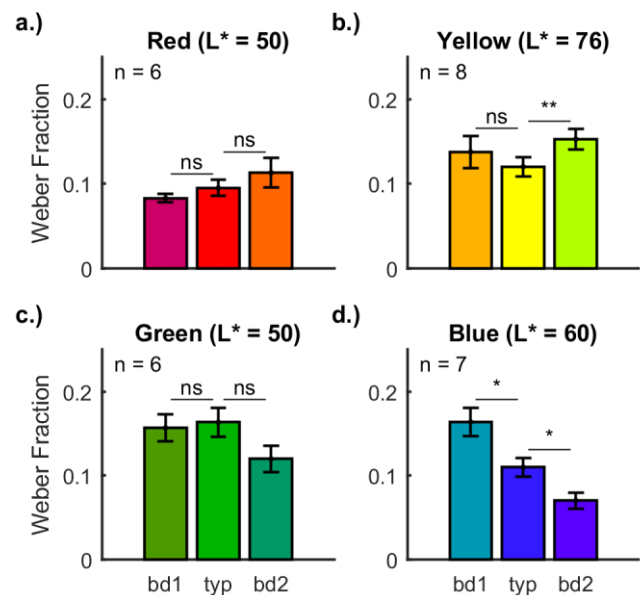
**Figure 5.** Threshold intensity plot (TVI) for observer f1 in Experiment 1. The x-axis represents CIELUV chroma of the test colors as radius in  $u^*v^*$ . JNDs (y-axis) are differences in  $u^*v^*$  radius between test and just-noticeable comparisons. Black circles correspond to measured JNDs; blue lines are fits to these JNDs with linear functions, the green curves fits with power functions. The red line is the extension of the blue line up to the visible gamut, along which JNDs were extrapolated for the main analyses. The panels in the first (a-c), second (d-f), third (g-i), and fourth (j-l) row show data for the red, yellow, green, and blue stimulus sets, respectively. The panels in the center column (b, e, h, k) depict results for typical hues, those on the left (a, d, g, j) and right (c, f, i, l) side show the JNDs for lower- and upper azimuth boundaries. Corresponding results for other observers may be found in Figures S4-S5 of the [Supplementary Material](#). Note that JNDs increase linearly as a function of test color radius (cf. results), in line with Weber-Fechner law (cf. discussion).

### Sensitivity

We then examined whether sensitivity is higher for typical than for boundary colors. **Figure 6** illustrates the Weber fractions (for exact numbers see column WF of [Table S4](#) in the [Supplementary Material](#)). The bars represent the Weber fractions (slope of linear TVI-plot). The lower the Weber fraction, the higher the sensitivity. So, if sensitivity for

colorfulness was particularly high for prototypes, the center bars in **Figure 6** were expected to be lower than the two other bars.

Only yellow (**Figure 6.b**) showed such a pattern. To compare measurements between boundary and typical hues, we computed a one-way repeated measurements ANOVA with the between-subjects factor hue (one typical, two boundary hues) for each category (red, yellow, green, blue) separately (cf. [Table S5](#) in the [Supplementary Material](#)). Results for red ( $F(2,10)=2.1$ ,  $p=0.17$ ) and green ( $F(2,10)=2.4$ ,  $p=0.14$ ) were not significant, which is likely due to the low number of observers ( $n=6$ ). Weber fractions differed significantly across hues in the yellow ( $F(2,14)=4.1$ ,  $p=0.04$ ) and blue category ( $F(2,12)=17.7$ ,  $p<0.001$ ). We calculated paired, two-tailed  $t$ -tests across participants to compare Weber fractions between the typical and each boundary hue. Following a Bonferroni correction for two  $t$ -tests, the significance level for these  $t$ -tests is  $\alpha = 0.025$ . Details on the  $t$ -tests are reported in [Table S6](#) of the [Supplementary Material](#). Sensitivity at typical yellow (**Figure 6.b**) differed significantly from sensitivity at the yellow-green boundary ( $t(7)=-4.8$ ,  $p=0.002$ ), but not from the one at the yellow-orange boundary ( $t(7)=-1.6$ ,  $p=0.15$ ). Sensitivity at typical blue differed significantly from the blue-green ( $t(6)=-3.0$ ,  $p<0.025$ ) and the blue-purple ( $t(6)=3.4$ ,  $p=0.02$ ) boundary; but the Weber Fraction for blue-purple was lower than the one for typical blue, which contradicts the predictions (**Figure 6.d**).



**Figure 6.** Sensitivity in Experiment 1. Panels a, b, c, and d refer to the red, yellow, green, and blue stimulus sets. The left (bd1) and right bar (bd2) in each graphic correspond to the lower- and upper-azimuth boundary of the category, respectively, the center (typ) bar to the typical hue. The lightness of the stimulus set ( $L^*$ ) is given in the titles. The y-axis represents Weber fraction. Colored bars show the Weber fractions averaged across participants with error bars indicating standard errors of mean. Symbols above bars



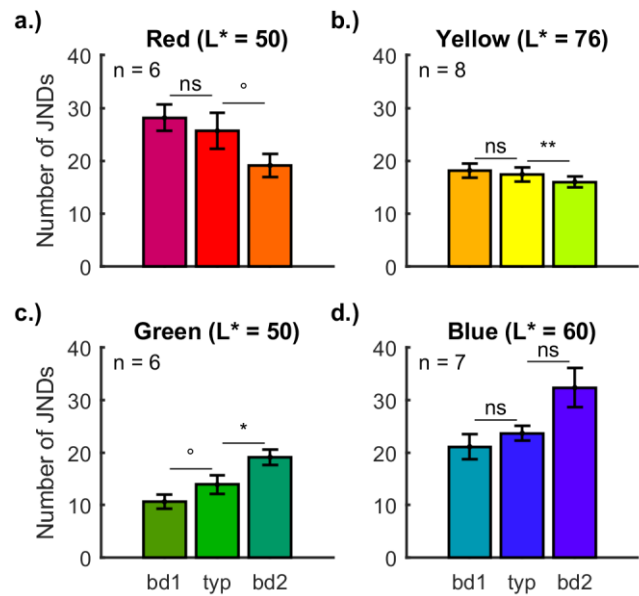
report results of t-tests after Bonferroni-correction. *Note that only for yellow were the center bars lower than the boundary bars.*

### Visible saturation

Finally, we tested whether visible saturation is higher for typical than for boundary colors. We calculated the CIELUV values for spectral lights at equal lightness as the stimuli in order to determine the visible gamut (thick black curves in **Figure 4**). We linearly interpolated JNDs up to the visible gamut (red lines in **Figure 5**). Then we calculated the cumulative number of JNDs up to the visible gamut, including the detection threshold (i.e. the threshold at the achromatic gray-point). The resulting visible saturation depends on the size of the detection threshold, the Weber Fraction and the visible gamut in the respective hue dimension, but it is independent of the scaling of the axes along which JNDs were measured (here the scaling of CIELUV radii). **Table S4** in the **Supplementary Material** provides detailed average specifications for these calculations. For the analysis below, these calculations have been done for each individual separately.

**Figure 7** shows the visible saturation for each hue direction and category. If prototypes can reach a particularly high visible saturation the center bars should be higher than both of the other bars. This was not the case for any of the categories. Repeated measures ANOVAs with the factor hue showed that the differences between colors were significant in all four categories (all  $p < 0.02$ ; cf. **Table S5** of the **Supplementary Material**). **Table S6** of the **Supplementary Material** provides details of the t-tests comparing visible saturation at each boundary hue to the typical hue. For red (**Figure 7.a**), none of the t-tests reached significance after Bonferroni correction ( $\alpha=0.025$ ), although the difference at the red-orange boundary was close to significance ( $t(5)=2.9$ ,  $p=0.03$ ). For yellow (**Figure 7.b**), t-tests showed that the number of JNDs at typical yellow were significantly higher than those at yellow-green ( $t(7)=4.3$ ,  $p=0.004$ ). For green (**Figure 7.c**), the difference at the green-blue boundary was significant ( $t(5)=-3.8$ ,  $p=0.01$ ), and the one at the green-yellow boundary just missed significance after Bonferroni correction ( $t(5)=2.8$ ,  $p=0.04$ ). For blue (**Figure 7.d**), t-tests were not significant (max.  $t(6) = -2.2$ , min.  $p = 0.07$ ).

In sum, the significant results for the green category clearly contradict the prediction of higher peaks at typical hues. The other patterns also seem to contradict this idea, but are difficult to interpret because of the lack of significance.



**Figure 7.** Visible saturation in Experiment 1. The y-axis represents the number of JNDs between the gray background and the visible gamut. Colored bars show the number of JNDs averaged across participants with error bars indicating standard errors of mean. Apart from that, format is as in **Figure 6**. *Note that for none of the categories were the center bars higher than the boundary bars.*

### Discussion

JNDs increased linearly with the CIELUV radius of the test color. Comparisons across hues contradicted the idea that prototypes involve higher sensitivity or higher visible saturation than boundary colors.

#### Linearity of JNDs

The empirical measurements of JNDs followed quite clearly a linear trend, as exemplified by the data of observer f1 in **Figure 5** and for other observers in **Figures S4-S5**. This was the case for all observers. A linear relationship between JNDs and radius has previously been observed in cone-opponent DKL-space (Krauskopf & Gegenfurtner, 1992). For isoluminant colors a line in CIELUV space corresponds to a line in cone-opponent space, and the nonlinearities in the scaling of the axes are rather small (cf. **Figure S1**). At isoluminance, distances along  $u^*$  are only rescaled in DKL-space. Nonlinearities only affect the  $v^*$  axis and are small. This explains why the JNDs measured in DKL space (Krauskopf & Gegenfurtner, 1992) and the JNDs measured in CIELUV space here both follow the Weber-Fechner law.

#### Statistics across participants

This experiment favored a large range of measurements for each single observer over a large sample of different observers. As a consequence, the statistical power for tests across participants was quite low. This is particularly true with respect to the variation of category prototypes and

unique hues across participants. Due to this variation, each participant was measured for a different hue, and the hues that corresponded to prototypes for one participant could well be close to the category boundary of another participant (**Figure S2**). Some of the results contradicted the idea of saturation peaks, but it seems advisable to obtain further evidence with larger samples that allow for tests with higher statistical power.

### **Global and local changes in saturation**

In this experiment, only 3 hues (one typical and two boundary hues) were compared in each stimulus set. For this reason, the results only allowed comparisons of typical and boundary hues in terms of their absolute size of sensitivity and visible saturation. However, it might be that sensitivity and visible saturation are modulated by other factors, which combine with effects of prototypes. In this case, discriminable saturation should peak locally around typical hues but not necessarily globally. For example, when considering the decrease of the visible saturation in the red category (**Figure 7.a**) as a global trend that combines with effects around prototypes, then the visible saturation would be higher for typical red than for the two boundaries together. Hence, the question arises whether there are local peaks at the prototypes that are partly covered by global modulations of JNDs and gamut (for a similar reasoning concerning JNDs for hue differences see Witzel & Gegenfurtner, 2013). In addition, we observed in the preliminary measurements of lightness and hue that at least some of the prototypes differed from unique hues (e.g. red), and that some hues might change with saturation (e.g. blue). It seems advisable to measure sensitivity and visible saturation across a larger range of hues in order to account for variations in hue and for global variations in sensitivity and visible saturation across hue.

## **Local peaks of saturation (Exp2)**

This second main experiment was designed to follow up the above results with larger samples of participants, and to examine the variation of saturation across hues at a more fine-grained resolution of hue. This allowed for examining local peaks in sensitivity and visible saturation, and to appreciate whether the difference between typical and unique red and green is important to the conclusions concerning the red and green stimulus sets. The above measurements had shown that JNDs very closely followed the Weber-Fechner law. For this reason, we measured JNDs in this experiment only at the adaptation point to establish the detection threshold (intercept) and at a high level of saturation to estimate the slope (Weber-Fraction) of the linear threshold-vs.-intensity function. With these measures, all JNDs between adapting gray-point and visible gamut may be linearly interpolated. This approach made it possible to measure a larger sample of hues across the categories and to test a large sample of participants. Apart from that, this

experiment aimed at answering the same questions as the above experiment: Does the sensitivity to saturation locally increase around typical and unique hues? And is visible saturation higher for typical and unique than for other hues?

## **Method**

### **Participants**

Overall, 28 observers participated in JND measurements, out of which 23 completed the measurements for red, and 20 for yellow, green, and blue. One observer in the yellow and one in the green stimulus set were remeasured at lower CIELUV chroma (65 instead of 70, and 30 instead of 35) because their JNDs were so large that comparison colors were out of gamut for the original test-colors. For the same reason, the dataset of two (of originally 22) observers for the yellow and of one observer (of originally 24) for the red stimulus set had to be excluded because these observers were not available for a re-measurement at lower CIELUV chroma. Observers in this experiment had also participated in the measurements of prototypes and unique hues.

### **Stimuli**

Test colors were sample from hue circles in CIELUV space. The radius of each hue circle (CIELUV chroma) was determined so that test colors were as saturated as possible while still allowing for comparison colors within the monitor gamut. As a result, the radius of the hue circle was set to 65 for red, 70 for yellow, 35 for green, and 40 for blue. The upper limit of the comparison colors was given by the monitor gamut. The minimum difference in CIELUV radius between the monitor gamut and the test colors was 19 for red, 20 for yellow, 16 for green, and 15 for blue, which is more than 1 JND above the test saturation.

### **Procedure**

The same 4 AFC task was used as in Experiment 1. No preliminary matching took place. Presentation time was 200ms (instead of 500ms as in Experiment 1) to further counteract the possibility of after-images. This slightly increased the size of JNDs. JNDs at and away from the adaptation points were measured separately. Increasing and decreasing staircases for each test hue were measured in one block, resulting in 10 blocks with 2 staircases each. The order of blocks was randomized.

## **Results**

**Figure 8** illustrates the aggregated measurements (typical hues and JNDs). **Figure 9** and **Figure 10** show the resulting Weber Fractions and visible saturation used for the main tests.

### **Typical hues and JNDs**

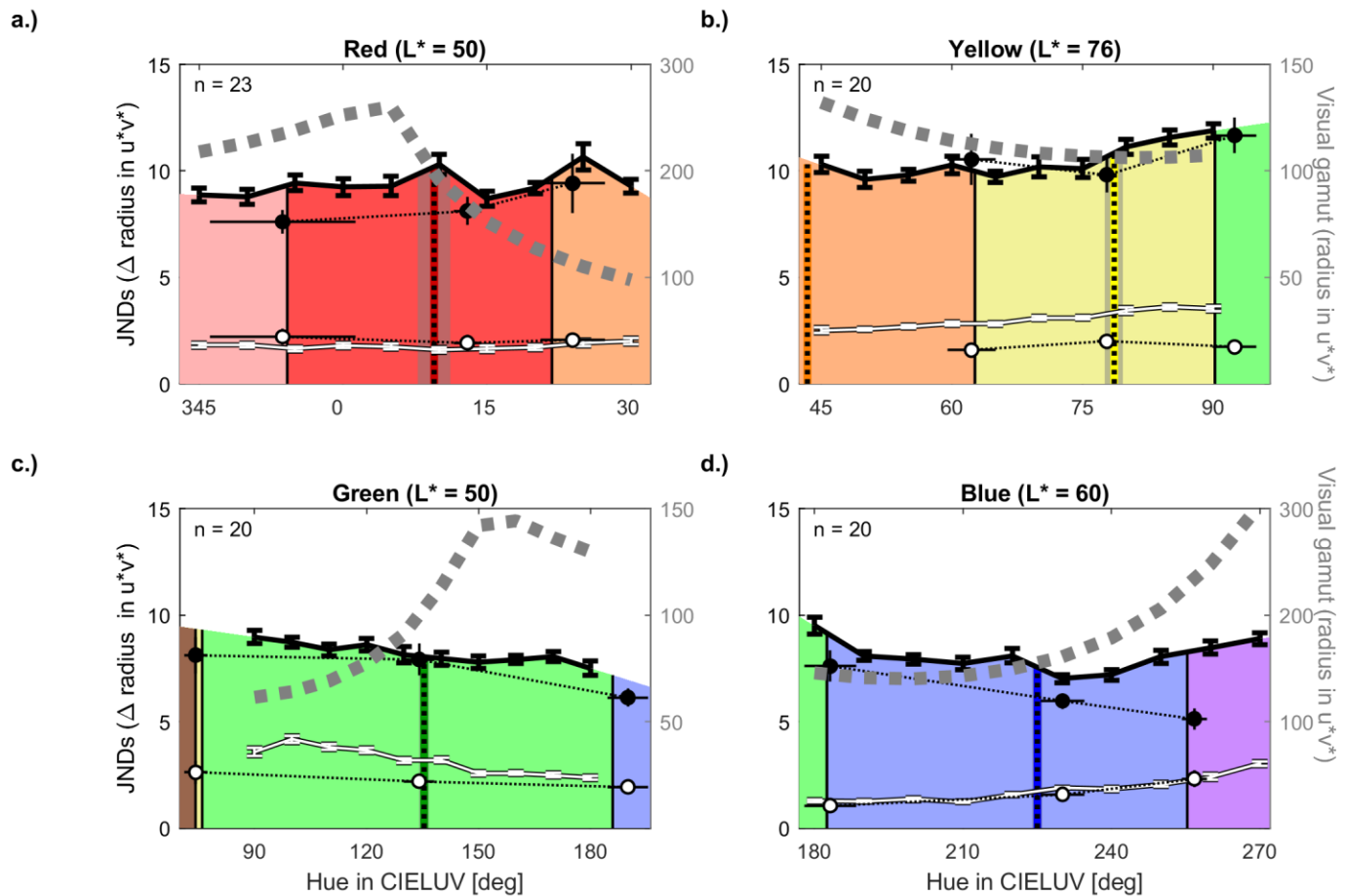
We lumped together typical and unique hues by averaging, for each participant separately, across all typical and unique hue adjustments. We will refer to these hues as *combined*

typical and unique hues (cf. “combined” in **Table 2**). The vertical dotted lines in **Figure 8** illustrate these combined typical and unique hues (see also **Figure 2** and **Figures 9-10**). As shown above, yellow at 78.6 deg and blue at 225.0 deg (**Figure 8.b-d**) are representative for both typical and unique hues. To simplify the presentation of results, we also focused on the combined measure of red (9.5 deg) and green (135.4 deg) in a first step of analysis (**Figure 8.a,c**). Where appropriate, we will consider the potential impact of differences between typical and unique red (17.5 vs. 8.2 deg) and green (130.6 vs. 136.7 deg).

As above, we discarded the first reversal-point and calculated JNDs as the average over the remaining four (2 up and 2 down) reversal-points. **Figure 8** shows the average JNDs across the 10 hues of each stimulus set. The black and the white curves correspond to JNDs for saturated colors away from the adaptation point and for colors at the adaptation point (detection thresholds), respectively. For cross-validation, the black and white disks show the results from Experiment 1, with the hue corresponding to the average across observers (cf. **Table S4**). Overall, the JNDs from Experiment 1 are quite similar to those obtained here; small differences may be due to the fact that the JNDs were measured for the individual typical and boundary hues,

which vary across observers and which may affect average JNDs due to asymmetries.

JNDs and detection thresholds follow global modulations across hues. We calculated correlations between average thresholds and azimuth to assess global modulations. For yellow (cf. black and white curves in **Figure 8.b**), JNDs ( $r(8)=0.83$ ,  $p=0.003$ ) and detection thresholds ( $r(8)=0.98$ ,  $p<0.001$ ) increase with ascending azimuth; for green (**Figure 8.c**) they decrease with azimuth ( $r(8)=-0.91$ ,  $p<0.001$ ;  $r(8)=-0.93$ ,  $p<0.001$ ). Not surprisingly then, JNDs and detection thresholds were correlated for yellow ( $r(8)=0.89$ ,  $p<0.001$ ) and green ( $r(8)=0.84$ ,  $p=0.002$ ). For red (**Figure 8.a**), JNDs and detection thresholds were generally rather flat, and there was no correlation between thresholds and azimuth (both  $p>0.30$ ) or between the two kinds of thresholds ( $r(8)=0.03$ ,  $p=0.93$ ). For blue (**Figure 8.d**), detection thresholds increased with azimuth ( $r(8)=0.92$ ,  $p<0.001$ ), but JNDs were neither correlated with azimuth ( $r(8)=-0.16$ ,  $p=0.65$ ) nor with detection thresholds ( $r(8)=0.12$ ,  $p=0.74$ ). Instead, JNDs seemed to follow a slightly U-shaped curve for bluish colors (black curve in **Figure 8.d**). These global trends are likely due to the inhomogeneity of CIELUV-space with respect to sensitivity.



**Figure 8.** JNDs in CIELUV (Experiment 2). The graphics show JNDs as the difference between test and comparison in CIELUV space along the left y-axis. The x-axis corresponds to hue in azimuth degree. The thick black curve above the colored areas refers to JNDs at the test saturation, the white curve shows detection thresholds at the adaptation point. Color areas indicate color categories, with vertical black lines being the average boundaries. For comparison the disks reflect the thresholds obtained in Experiment 1 at the adaptation point (white) and at the CIELUV radius of the test colors used in this Experiment 2 (black). Dark dotted vertical lines show the combined typical and unique hue (see main text for details), and the gray band around them corresponds to the standard error of mean. The right y-axis shows radius in CIELUV, and corresponds to the dotted curve which illustrates the visible gamut in CIELUV. The detection thresholds (white curve), the gamut (gray curve), and the Weber fractions in Figure 9 explain the patterns of visible saturation shown in Figure 10.

### Weber-Fractions

Weber-Fractions were calculated as the slope of JNDs (ratio between JNDs and CIELUV chroma) taking detection thresholds (white curve in Figure 8) as the intercept. Figure 9 illustrates Weber Fractions across hues (black curves). If sensitivity to saturation was particularly high for typical and unique hues, Weber Fractions should locally decrease towards typical and unique hues.

As for JNDs, some global trends are visible. In particular, Weber-Fractions for bluish colors (Figure 9.d), strongly decrease with azimuth, resulting in a negative correlation with azimuth ( $r(8) = -0.73$ ,  $p = 0.02$ ). This decrease of Weber fractions may be explained by the increase of detection thresholds (cf. above). Weber fractions for yellowish hues (Figure 9.b) show a close to significant tendency to increase with azimuth ( $r(8) = 0.59$ ,  $p = 0.07$ ); but the overall pattern was

slightly U-shaped, rather than linear. Weber fractions for red (Figure 9.a) and green (Figure 9.c) seem rather flat with local troughs and peaks, and hence no correlations with azimuth were observed (both  $p > 0.19$ ).

We also calculated repeated measures analyses of variance (RMAOVs) with the factor hue to test for differences across the 10 hues (details in Table S8). Weber fractions differed significantly for red ( $F(9,198) = 2.1$ ,  $p = 0.03$ ; cf. Figure 9.a) and blue ( $F(9,171) = 7.0$ ,  $p < 0.001$ ; cf. Figure 9.b); but tests were not significant for yellow and green (both  $p > 0.37$ ).

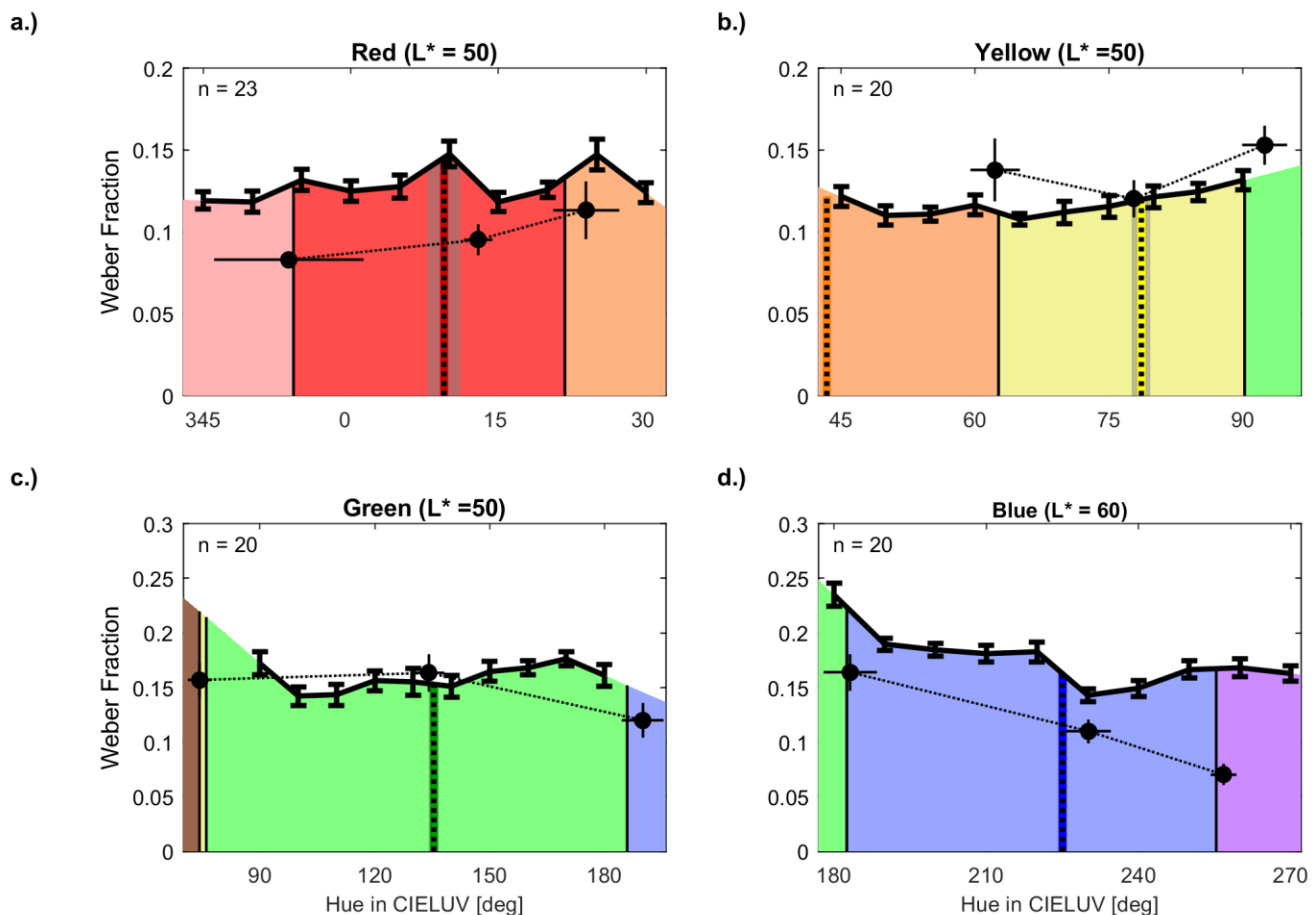
We then examined local modulations of sensitivity that are specific to typical and unique hues. We conceived *locality tests* to account for global trends when testing for local modulations. They test whether Weber-Fractions are lower around the typical and unique hues than predicted by the global trend of Weber-Fractions. To account for such global



trends, we calculated the regression line across the ten test hues of each stimulus set for each participant. The regression line provides a “predicted” Weber-Fraction for typical and unique hues, if Weber-Fractions just followed the overall trend without any focal color effect. To determine the Weber Fraction at the typical and unique hue, we linearly interpolated the Weber-Fractions of the two test hues adjacent to the typical and unique hue. This was done for each individual observer separately. Then we tested with paired, two-tailed t-tests across participants whether the Weber-Fractions at the typical and unique hues were below the regression line. If there were local peaks of sensitivity around the typical and unique hues, Weber-Fractions at the typical and unique hues should be lower than those predicted by the respective regression lines across hues.

However, none of the locality tests were significant (all  $p > 0.16$ ). Detailed results of these tests are provided by Figure

S6 and in Table S10 of the [Supplementary Material](#). The absence of a significant pattern in the red category could be due to the fact that typical red is shifted towards pink compared to unique red, which is around an azimuth of 17.5 deg (Figure 9.a). If we consider typical and unique red separately, unique red coincides with the local dip of Weber Fractions between the hues at 15 and 20 deg. For blue, there seems to be a decrease of Weber fractions towards typical blue in the aggregated data in Figure 9.d. Sixteen of 20 observers yielded Weber fractions at the typical and unique hue below the regression line (cf. Figure S6.d and Table S10). However, two observers with slightly greenish typical blue produced Weber fractions at the typical and unique hue that were very much above the regression line, hence undermining a significant result for blue. Weber Fractions in the yellow and green category seem to clearly contradict the idea that sensitivity increases specifically towards typical yellow and green (Figure 9.b-c).



**Figure 9.** Weber fractions in Experiment 2. The y-axis and the black curve represent Weber-Fractions. The black disks reproduce Weber fractions from Experiment 1. Apart from that, format is as in Figure 8. *Note that Weber Fractions tend to increase towards typical and unique red and decrease towards typical and unique blue, but there is no specific pattern at typical and unique yellow and green.*

### Visible saturation

We used detection thresholds (intercept, white curve in Figure 8) and Weber-Fractions (slope, black curve in Figure 9) to linearly extrapolate JNDs to the visible gamut (dotted

gray curve in Figure 8). Then, JNDs between the adaptation point and the visible gamut were summed up across the visible gamut to determine the visible saturation. Figure 10 illustrates the visible saturation for each hue (black curve).

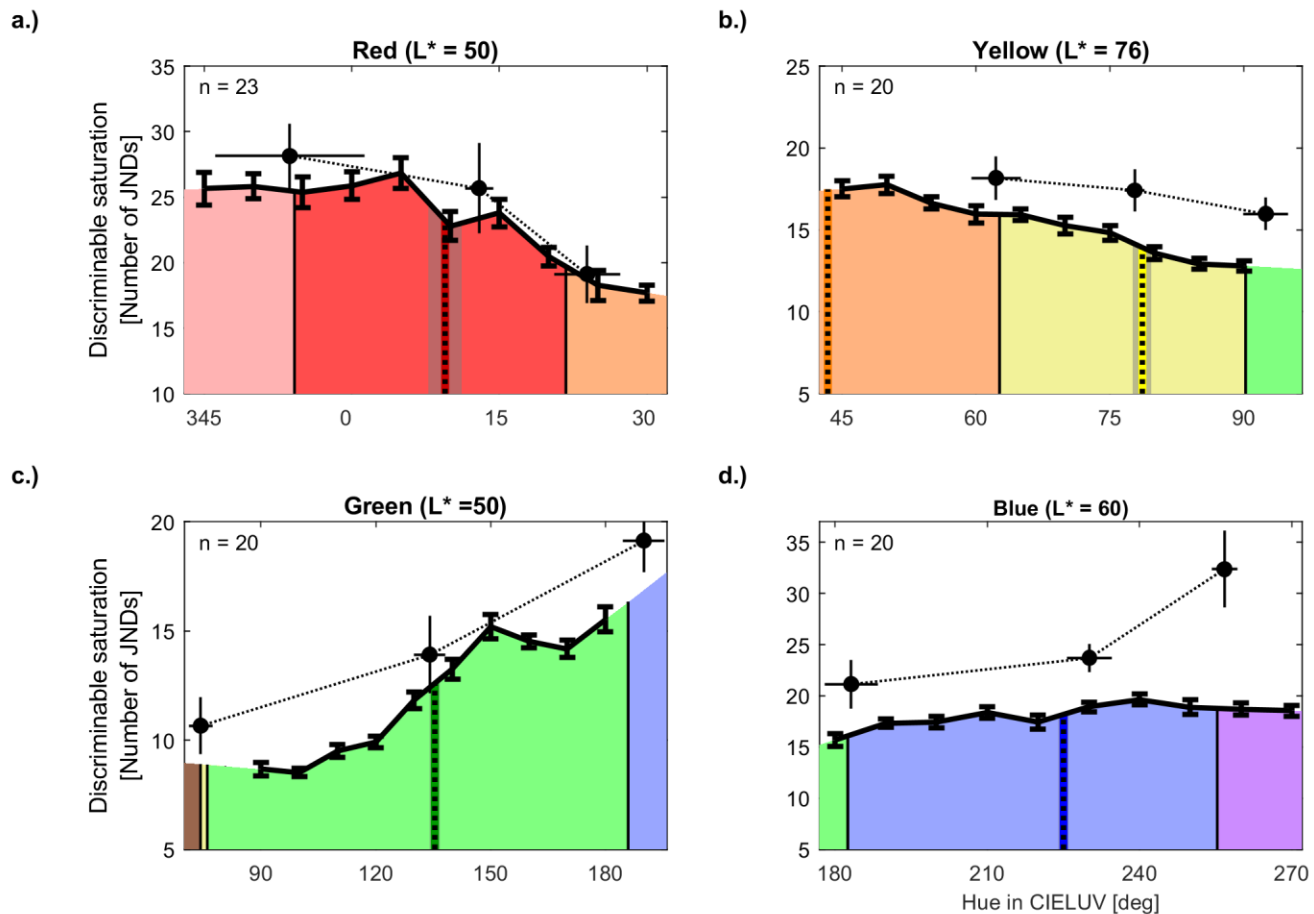
The higher the black curves in **Figure 10**, the more discriminable levels of saturation fit between the adapting gray-point and the visible gamut for a given hue. For comparison, the black disks reproduce the results from Experiment 1 (**Figure 7**). The higher absolute size of the visible saturation in Experiment 1 compared to Experiment 2 reflects the impact of the lower JNDs in Experiment 1 (cf. **Figure 8**).

There were strong differences of visible saturation across hue ranges. Red and pink colors reach over 25 JNDs away from the adapting gray point (**Figure 10.a**), while brown-green colors have a visible saturation of less than 10 JNDs (left side of **Figure 10.c**). For all four categories, visible saturation also strongly varied across hues, as shown by significant RMAOVs with the factor hue (all  $F > 3.6$ , all  $p < 0.001$ ; cf. **Table S8**). Visible saturation shows also very clear global trends that can be captured by correlations across hue (cf. **Table S9**). Visible saturation decreases with azimuth for red ( $r(8) = -0.87$ ,  $p < 0.001$ ) and yellow hues ( $r(8) = -0.98$ ,  $p < 0.001$ ), and increases for green ( $r(8) = 0.95$ ,  $p < 0.001$ ) and blue hues ( $r(8) = 0.77$ ,  $p = 0.009$ ).

If focal colors reach higher levels of visible saturation the black curve should have local peaks around the typical and unique hues (vertical dotted lines). We tested this pattern using the locality tests described above (section on Weber-

Fractions). To account for global trends, we calculated, for each observer, the regression line across the ten test hues of a category. Then, we determined with a paired, two-tailed  $t$ -test whether the number of discriminable colors at the typical and unique hue was significantly above the value predicted by the regression line. None of the four locality tests was significant (all  $p > 0.11$ ). Detailed results are provided by **Figure S7** and **Table S10** of the [Supplementary Material](#).

The results for red (**Figure 10.a**) are again complicated by the potential difference between typical and unique red (17.2 vs. 8.2 deg, **Figure 2.a**). However, the patterns of yellow, green, and blue clearly contradicted the idea that typical and unique hues coincide with peaks of visible saturation. For yellow and blue (**Figure 10.b&d**), visible saturation almost completely followed a global trend, and did not show local changes. For green (**Figure 10.c**), the local peak at 150 deg did not coincide with the typical and unique hue at 135 deg. The comparatively large difference of the local peaks of visible saturation from combined typical and unique green may not be attributed to the much smaller, potential difference between typical and unique green (131.0 vs. 136.7 deg, **Figure 2.c**). Taken together, these results suggest that the variation of visible saturation across hues is independent of category prototypes and unique hues.



**Figure 10.** Visible saturation in Experiment 2. The graphics show the number of JNDs away from the adapting gray point along the y-axis. The thick black curve shows the number of JNDs between gray and visible gamut (visible saturation). Apart from that format is as in Figure 8. *Note that visible saturation strongly changes across hues, but at least for yellow, green and blue it does not peak at typical and unique hues.*

## Discussion

None of the four hue ranges showed clear peaks in sensitivity around typical and unique hues except maybe for blue (Figure 9). Although visible saturation strongly varied across hues, at least yellow, green, and blue clearly contradicted the idea that visible saturation peaks around typical and unique hues (Figure 10). Results largely agreed with those found in Experiment 1 (cf. circles in Figures 8-10), except for the blue-purple boundary hue (panel d in Figures 8-10). These results have implications for the main question concerning the perceptual salience of focal colors, for the use of maximally saturated Munsell chips in color naming research, and for theories of unique hues.

### Perceptual salience of focal colors

Both experiments of this study contradicted the idea that colors with typical and unique hues are perceptually salient if we define salience in terms of discriminable saturation, i.e. the ability to discriminate different levels of saturation. If such colors were perceptually salient, we would expect all, or

at least most, of the colors to coincide with peaks in sensitivity or visible saturation; but this was not the case. There was some uncertainty concerning the precise hue direction for red, green, and blue due to overall variability (red), differences between unique and typical hues (red and green), and changes with saturation (blue). In addition, a trend towards lower Weber fractions was visible for blue, but might have missed significance due to low statistical power. However, if peaks of saturation were a general feature of typical and unique hues, they should appear for all four categories. This is clearly not the case since most categories follow global modulations of sensitivity (Weber fractions) and visible saturation (Figure 9 and Figure 10).

It is theoretically possible that the subjective appearance of saturation (i.e. subjective saturation, cf. Introduction) dissociates from discriminable saturation. For example, an equally discriminable difference might subjectively appear larger for some than for other hues. Defining perceptual salience in terms of subjective saturation rather than discriminable saturation might be a valid alternative approach to test the present research question. In a

companion study (Witzel & Franklin, 2014), we compared subjective saturation to discriminable saturation (based on the aggregated data of JNDs from the present Experiment 2). In that study, subjective saturation was measured by matching the saturation of a comparison color to another color with fixed saturation, the test color. To represent results in terms of discriminable saturation, we used a polynomial fit for the data on discriminable saturation to avoid noise in JND measurements to interfere with the examination of subjective saturation (cf. **Figure S8** in the [Supplementary Material](#)). When comparing discriminable saturation between the two studies, note that the discriminable saturation reported here (between 7 and 40 JNDs in **Figure 10**) corrects for an error in scaling due to which the absolute size of discriminable saturation is overestimated in the companion study (up to 70 JNDs for red; cf. Figures 3-4 in Witzel and Franklin, 2014). This scaling error was irrelevant for the conclusions of that study, which were based on relative differences across hue, not absolute size. The results showed that typical and unique hues did not appear more colorful than other hues with equivalent discriminable saturation (see Figure 3 in Witzel & Franklin, 2014). The study also showed that maximum subjective saturation for spectral colors at the visible gamut does not peak around prototypes (Figure 4 in Witzel & Franklin, 2014).

The data of the present study allow us to re-evaluate and extend the results from the companion study. The companion study examined subjective saturation for equal levels of (average) discriminable saturation (Figure 3-4 in Witzel & Franklin, 2014). **Figure 11** allows for a complementary approach that assesses discriminable saturation for equal levels of subjective saturation. For this purpose, matches of subjective saturation from Witzel and Franklin (2014) were averaged for a given comparison hue and across all test colors. These average matches correspond to colors that subjectively appear to be equally saturated. If subjective saturation was exactly the same as discriminable saturation the black curves in **Figure 11** would be constant across hues, i.e. horizontal lines.

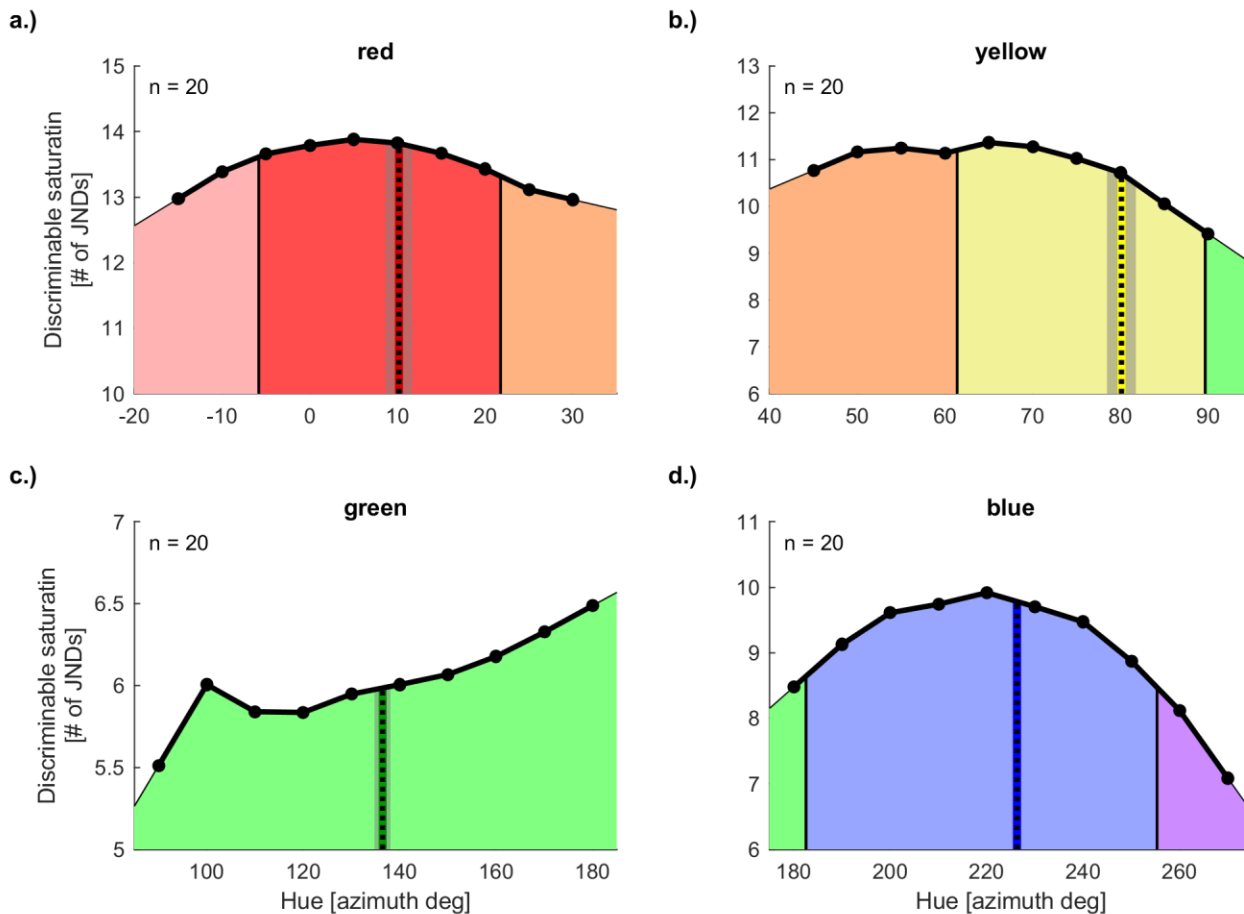
Instead, subjective saturation varies systematically across hues. This indicates that subjective saturation is not completely determined by discriminable saturation. For red and blue, and maybe for yellow, curves of equal subjective saturation have global peaks that seem to roughly coincide with typical and unique hues. These patterns imply that higher levels of discriminable saturation are necessary to reach equal amounts of subjective saturation at prototypes

and unique hues. This observation contradicts the idea of high subjective saturation around typical and unique hues, and hence reconfirms our previous observations on subjective saturation based on different analyses (Witzel & Franklin, 2014).

Taken together, the findings from the present and the companion study (Witzel & Franklin, 2014) refute the idea that the typical and unique red, yellow, green, and blue are more colorful, and hence more salient, than other hues. Consequently, these colors are not focal in the sense that they have particularly high levels of colorfulness and perceptual salience. Other approaches also contradicted the idea that prototypes and unique hues have particular perceptual characteristics that qualify them as focal colors or perceptual anchors. Hue discrimination does not show patterns specific to typical and unique hues (Witzel & Gegenfurtner, 2013; Witzel & Gegenfurtner, 2016; Witzel & Gegenfurtner, 2018a). Perceptual measures of color constancy do not peak around typical hues when controlling for saturation (Witzel, van Alphen, Godau, & O'Regan, 2016; Weiss, Witzel, & Gegenfurtner, 2017). Hue scaling can be accomplished using elementary hues other than unique red, green, yellow, and blue (Bosten & Boehm, 2014), and there is evidence against the idea that unique hues are more reliable (less variable) than intermediate hues (Bosten & Lawrence-Owen, 2014). Still another study observed that unique hues did not yield higher salience in a visual search task (Wool et al., 2015).

All those observations strongly undermine the idea that category prototypes and unique hues have particular perceptual properties and that they may act as perceptual anchors for color appearance and color naming (for further discussion see also Witzel, 2018a, 2018b; Witzel & Gegenfurtner, 2018b). Alternatively, category prototypes may be the result of, rather than the cause of color categorization (cf. Lindsey & Brown, 2019). It has been suggested that similarities of color categorization across cultures is optimized to communicate efficiently about the irregularly shaped perceptual color space (Abbott et al, 2016; Zaslavsky et al., 2018; Zaslavsky et al., 2019). Evidence for this idea is based on the irregular distribution of maximally saturated Munsell chips and naming data from the World color survey. Others suggested that cross-cultural tendencies of categories and their prototypes might be related to objects and materials in the natural environment (Gibson et al., 2017; Siuda-Krzywicka et al., 2019; Witzel, 2018a; Witzel & Gegenfurtner, 2018; Yendrikhovskij, 2001; but see also Zaslavsky et al., 2019).





**Figure 11.** Subjective saturation as a function of discriminable saturation. Curves show equal levels of subjective saturation according to the matches measured by Witzel & Franklin (2014). The x-axis represents hue as CIELUV azimuth in degree, the y-axis the discriminable saturation as the number of JNDs away from the adaptation point. Discriminable saturation is estimated based on aggregated JNDs fitted by a second-order polynomial to discount noise (cf. Figure S8 in the [Supplementary Material](#)). Apart from that, format as in Figure 8. *If subjective saturation was equivalent to discriminable saturation the curves would be flat lines.*

### Maximally saturated Munsell chips

Our results also imply that the coincidence of peaks in Munsell Chroma around category prototypes and unique hues is a peculiarity of the Munsell system, rather than a characteristic of color vision. The reason why Munsell chips provide higher levels of saturation for prototypical red, yellow, green, and blue may be the choice of pigments used to produce Munsell chips (Witzel et al., 2015; Witzel, 2018b). In particular, pigments with high chroma might coincide with English prototypes in the Munsell system. For example, the red pigments (sulphuret of mercury) were much more saturated than the green-blue pigments (sesquioxide of chromium) in the original Munsell system (Munsell, 1912).

Previous observations of unique perceptual characteristics (Regier et al., 2007; Philipona & O'Regan, 2006; Kuehni et al., 2010; Logvinenko, 2012; Logvinenko & Beattie, 2011; Vazquez-Corral et al., 2012; Lindsey et al., 2015) may well be explained by those peculiarities of the stimulus sample (Witzel, 2018a, 2018b; Witzel et al., 2015). Most

importantly, the original idea of focal colors was introduced to describe universals in color naming across languages (Berlin & Kay, 1969; Regier et al., 2005; Rosch Heider, 1972). It has previously been shown that color categorization depends on the variation of saturation in the set of maximally saturated Munsell chips (Witzel, 2016; Lindsey, Brown, Brainard, & Apicella, 2016; Witzel, 2018b, 2019). In particular, category membership and prototype choices are correlated with the saturation of the maximally saturated Munsell chips (for review see Witzel, 2018a, 2018b). In all studies supporting focal color salience, saturation was determined by Munsell chroma, assuming that Munsell chroma is roughly representative of perceived chroma. The precise measures of discriminable saturation in the present study allow for re-assessing the relationship between prototype choices and perceived saturation.

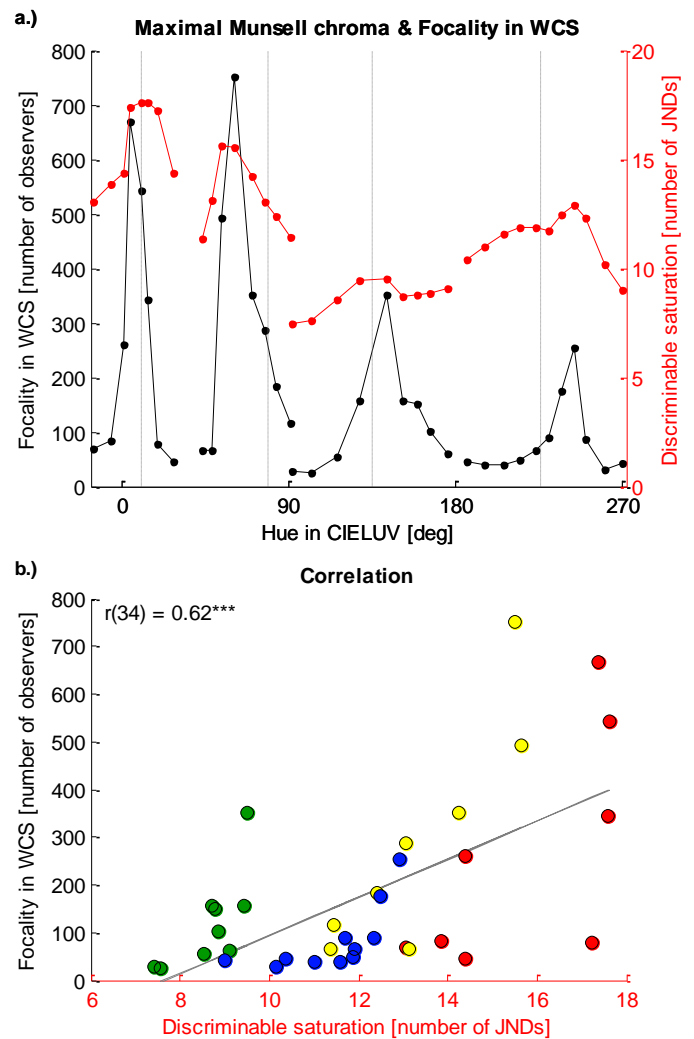
For this purpose, we identified the Munsell chips in the World Color Survey (Figure 1 in Regier et al., 2005) that correspond to the colors for which we had measured discriminable saturation. We represented Munsell chips in

CIELUV under standard illuminant C ( $xyY = [0.31006, 0.31616, 100]$ ). We then identified those Munsell chips whose CIELUV hues corresponded to the hues of our test colors and which had the most typical lightness in the World Color Survey. The Munsell Values of these chips were 4 for red with  $L^* = 41.2$ , 8 for yellow with  $L^* = 81.3$ , and 5 for green and blue with  $L^* = 51.6$  (cf. Figure 2 in Regier et al., 2005 or Figure 4.b in Witzel et al., 2015). Note that the lightness values of those prototypes were slightly different from the typical lightness we had measured ( $L^*=50$  for red and green, 76 for yellow, and 60 for blue). The reason for this discrepancy might be that  $L^*$  only approximately models lightness adaptation, and hence may not completely capture the lightness of reflectances.

We calculated discriminable saturation for the selection of maximally saturated Munsell chips, assuming that the JNDs are roughly similar despite the small differences in lightness. The resulting discriminable saturation for those Munsell chips is shown by the red curve in **Figure 12.a**. The black curve represents “focality”, i.e. the frequency of prototype choices across 110 languages in the World Color Survey for those Munsell chips (Regier et al., 2005). WCS prototype choices for red, yellow, green and blue coincide with local peaks of discriminable saturation. **Figure 12.b** illustrates the correlation between prototype choices and discriminable saturation, which was positive and highly significant ( $r(34)=0.62$ ,  $p<0.001$ ).

When saturation is determined as Munsell Chroma, CIELUV, and CIELAB radius the correlations are higher ( $r(34) = 0.79$ ,  $r(34) = 0.74$ , and  $r(34) = 0.62$ , respectively) than those with discriminable saturation (cf. **Figure S13** of the [Supplementary Material](#)). This seems to be mainly due to the saturation of green hues, which had a comparatively low discriminable saturation (green dots in **Figure 12.b**). This misalignment of the discriminable saturation of green hues with the other stimulus sets might be due to the discrepancies between the lightness of our test colors and the Munsell chips. However, the relationship between prototype choices and discriminable saturation seems to exist within each hue range: Observers tend to choose the most saturated colors in a hue range as prototypes.

These observations clearly confirm our earlier observations of a relationship between cross-cultural color categorisation and the variation of saturation in the stimulus set (Witzel et al., 2015; Witzel, 2016; Witzel, 2018b). Hence, the observed cross-cultural patterns in color categorisation may well be due to the unequal distribution of saturation and chroma in those stimulus samples (Berlin & Kay, 1969; Rosch Heider, 1972; Webster et al., 2002; Kay & Regier, 2003; Regier et al., 2005; Regier et al., 2007; Lindsey & Brown, 2006, 2009; Lindsey et al., 2015). These additional findings highlight the importance of controlling saturation and chroma in color naming research, and of reproducing cross-cultural patterns that have been found with maximally saturated Munsell chips.



**Figure 12.** Discriminable saturation and “focality” in the World Color Survey (WCS). Data points in this figure correspond to 36 maximally saturated Munsell chips, which is a subset of the stimuli used in the WCS (for details on the identification of these Munsell chips see main text). Panel a shows the variation of prototype choices (“focality”) in the WCS (black curve, left black axis) and the variation of discriminable saturation (red curve) across hues. The x-axis indicates hue as CIELUV azimuth in degree. The vertical dotted lines display typical and unique hues as measured in the present study: From left to right, red, yellow, green, and blue; but note that differences between the typical hues from our study and the hues from the WCS prototype choices can result from deficiencies of the CIELUV model of adaptation. Panel b provides a scatter plot that illustrates the correlation between WCS focality and discriminable saturation. Figure S9 of the supplementary material provides corresponding graphics for saturation measured as Munsell Chroma, CIELUV and CIELAB radius. *Note that the maxima of the Munsell chips (red curve in panel a) roughly coincide with the peaks of prototype choices (black curve), resulting in a correlation between the curves (panel b).*

### Mixture of unique hues

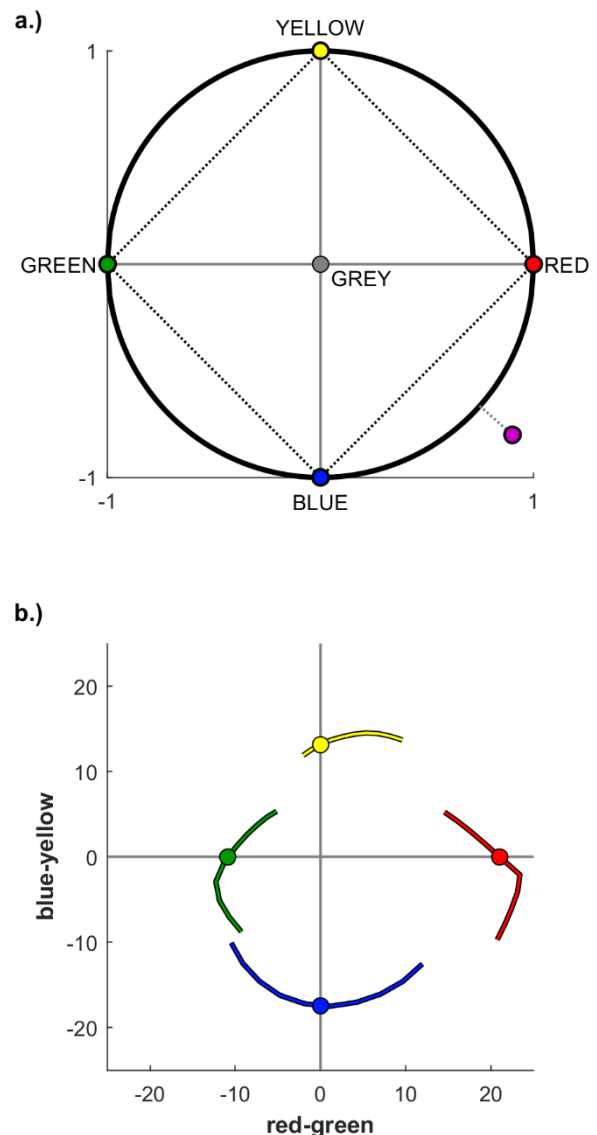
The observation that some intermediate hues can reach higher levels of visible saturation than unique hues has still further implications for the concept of unique hues. According to the original idea of unique hues, the opponent color pairs (red-green, blue-yellow, and black-white) constitute the poles of the 3 axes of a color appearance space that defines all apparent colors. The idea of such a space is illustrated by **Figure 13.a** (cf. Figure 2 in Jameson, 2010; Figure 2.a in Valberg, 2001). This idea is implemented for example in the Natural Color System (Hård, Sivik, & Tonnquist, 1996). In this space, the mixture of chromatic and achromatic unique hues, either along the hue circle (solid circle in **Figure 13.a**) or in chromaticity (dotted lines in **Figure 13.a**) cannot produce an intermediate color with a higher saturation (e.g. the purple disk in **Figure 13.a**) than the respective unique hues (e.g. the red and blue disks in **Figure 13.a**).

**Figure 13.b** reproduces the results from **Figure 10** in Cartesian coordinates. Hues have been rotated so that the unique hues constitute the four cardinal directions of a color-opponent space. In contrast to **Figure 10**, the data of **Figure 13.b** is based on the JND averages across observers. In order to discount for local variations, the aggregated JNDs have been smoothed across hues with a second-order polynomial function (cf. **Figure S8** in the [Supplementary Material](#)). Since lightness was the same for unique red and unique green ( $L^*=50$ ), but different for unique yellow ( $L^*=76$ ) and blue ( $L^*=60$ ), the plane of the unique hue axes in **Figure 13.b** does not correspond to a plane in CIELUV, but is slanted around the green-red axis when lightness is represented by  $L^*$ .

According to our estimations of visible saturation, hues on this plane in **Figure 13.b** do not form a circle like the one shown in the ideal model of **Figure 13.a**. Most importantly, bluish red, reddish yellow and bluish green show local protrusions in visible saturation. These are difficult to obtain through simple mixtures, i.e. direct linear or polar transitions between two adjacent unique hues. Other transitions are theoretically possible for reddish yellow and bluish green since their visible saturation is smaller than the visible saturation of at least one of the adjacent unique hues (i.e. blue and red, respectively). However, the visible saturation of bluish red is higher than both the visible saturation of unique red and of unique blue. Unique black and white may desaturate colors and produce all the colors with lower saturation than unique red, yellow, green, and blue. However, intermediate colors with a higher visible saturation than maximally saturated unique hues cannot result from a mixture of unique hues in such a simple color opponent space. One may wonder where this additional color bit of saturation comes from given that unique hues cannot have such a high level of visible saturation.

This observation undermines the idea that all intermediate colors may be produced by mixing unique hues. It joins

those studies that questioned the elementary nature of unique hues in color mixture and color appearance (Bosten & Boehm, 2014; Bosten & Lawrance-Owen, 2014; Wool et al., 2015). In this light, it seems particularly important to account for the role of saturation in future studies on unique hues and color appearance (for further discussion, see Witzel, 2018b).



**Figure 13.** Unique hues and saturation. The graphic illustrates an ideal representation of color appearance (cf. Figure 2.a in Valberg, 2001). The gray disc in the center represents achromatic (i.e. neutral) gray. Pure red, yellow, green, and blue correspond to unique, unmixed hues, and the distance from gray (i.e. the radius) indicates perceived saturation. The black circle illustrates mixtures of red, yellow, green, and blue at maximum saturation, the dotted square represents mixtures of maximally saturated unique hues in chromaticity. The purple disk illustrates the idea of

a purplish color that is more saturated than the maximal saturation of red, yellow, green, and blue, and hence is located outside the black circle. Panel b shows such an opponent space constructed based on our measurements of discriminable saturation. CIELUV hues are rotated so that the axes of this space coincide with unique hues. The distance to the origin reflects the discriminable saturation (Figure 10), calculated based on aggregated JNDs and smoothened with a polynomial fit. The four unique hues, and hence the four cardinal directions have different lightness levels ( $L^*$ ) in CIELUV. The colored curves indicate visible saturation, i.e. the maximum possible discriminable saturation for the respective hue directions. *Note that the four unique hues do not lie on a circle, and that intermediate hues (bluish red, reddish yellow, bluish green) have larger visible saturation than would be compatible with a simple transition between unique hues.*

## Conclusion

We investigated whether the typical and unique red, yellow, green, and blue are the most colorful colors. We defined colorfulness as “discriminable saturation”, which is the number of JNDs away from the adapting gray-point. The evidence from both experiments contradicted this idea. A companion study (Witzel & Franklin, 2014) assessed subjective saturation and refuted the idea that typical and unique hues have higher subjective saturation than intermediate hues. Since prototypes do not feature particularly high levels of chroma and saturation, the high chroma around typical red, yellow, green, and blue in the Munsell system does not reflect a particularity of the visual system, but simply a peculiarity of the Munsell system. Previous observations in support of the “focality” of typical and unique red, yellow, green, and blue found with maximally saturated Munsell might well be an artifact of the particular stimulus choice. Our findings also raise the question of how unique hues can be mixed to produce the color appearance of all other perceivable colors, given that they cannot attain the degree of colorfulness of many intermediate (non-unique) hues. Finally, we also observed a discrepancy between our measures of discriminable and subjective saturation. The rich dataset of saturation measures provided in this study may help to further evaluate and improve color spaces and to clarify the relationship between discrimination and subjective appearance of colorfulness in future studies.

## Acknowledgements

This research was supported by a German Academic Exchange Service (DAAD) postdoctoral fellowship to CW, a European Research Council funded project (“Categories”, Ref. 283605) to A.F, and a ERC Advanced Grant FEEL No. 323674 to J. Kevin O’Regan and by grant “Cardinal Mechanisms of Perception” No SFB TRR 135 from the Deutsche Forschungsgemeinschaft. We thank Ying Chen for considerable contributions to data collection.

## References

- Abbott, J. T., Griffiths, T. L., & Regier, T. (2016). Focal colors across languages are representative members of color categories. *Proceedings of the National Academy of Sciences*, 113(40), 11178-11183. doi:10.1073/pnas.1513298113
- Abramov, I., & Gordon, J. (1994). Color appearance: on seeing red - or yellow, or green, or blue. *Annual review of psychology*, 45, 451-485. doi: 10.1146/annurev.ps.45.020194.002315
- Berlin, B., & Kay, P. (1969). *Basic color terms: their universality and evolution*. Berkeley, CA: University of California Press.
- Bolton, R. (1978). Black, White, and Red All Over: The Riddle of Color Term Salience. *Ethnology*, 17(3), 287-311.
- Bosten, J. M., & Boehm, A. E. (2014). Empirical evidence for unique hues? *Journal of the Optical Society of America. A, Optics, image science, and vision*, 31(4), A385-393. doi: 10.1364/JOSAA.31.00A385
- Bosten, J. M., & Lawrance-Owen, A. J. (2014). No difference in variability of unique hue selections and binary hue selections. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 31(4), A357-364. doi: 10.1364/JOSAA.31.00A357
- Boynton, R. M., MacLaury, R. E., & Uchikawa, K. (1989). Centroids of color categories compared by two methods. *Color Research & Application*, 14, 6-15.
- Boynton, R. M., & Olson, C. X. (1990). Salience of chromatic basic color terms confirmed by three measures. *Vision Research*, 30(9), 1311-1317. doi: 0042-6989(90)90005-6 [pii]
- Brown, R. W., & Lenneberg, E. H. (1954). A study in language and cognition. *Journal of Abnormal and Social Psychology*, 49(3), 454-462.
- Burns, S. A., Elsner, A. E., Pokorny, J., & Smith, V. C. (1984). The Abney effect: chromaticity coordinates of unique and other constant hues. *Vision Research*, 24(5), 479-489.
- CIE. (1932). *Commission Internationale de l'Eclairage Proceedings, 1931*. Cambridge: Cambridge University Press.
- Collier, G. A. (1973). Review of "Basic Color Terms: Their Universality and Evolution". *Language*, 49(1), 245-248.
- Collier, G. A., Dorflinger, G. K., Gulick, T. A., Johnson, D. L., McCorkle, C., Meyer, M. A., . . . Yip, L. (1976). Further Evidence for Universal Color Categories. *Language*, 52(4), 884-890.



- Dzhafarov, E. N., & Colonius, H. (2011). The Fechnerian Idea. *The American Journal of Psychology*, 124(2), 127-140. doi: 10.5406/amerjpsyc.124.2.0127
- Fairchild, M. D. (2005). *Colour appearance models* (2. ed.). Hoboken, Nj: Wiley.
- Gibson, E., Furell, R., Jara-Ettinger, J., Mahowald, K., Bergen, L., Ratnasingam, S., . . . Conway, B. R. (2017). Color naming across languages reflects color use. *Proceedings of the National Academy of Sciences*. doi: 10.1073/pnas.1619666114
- Hansen, T., Walter, S., & Gegenfurtner, K. R. (2007). Effects of spatial and temporal context on color categories and color constancy. *Journal of Vision*, 7(4), 1-15.
- Hård, A., Sivik, L., & Tonnquist, G. (1996). NCS, natural color system - From concept to research and applications. Part I. *Color Research & Application*, 21(3), 180-205.
- Hays, D. G., Margolis, E., Naroll, R., & Perkins, D. R. (1972). Color Term Salience. *American Anthropologist (New Series)*, 74(5), 1107-1121.
- Hering, E. (1964 [1878]). *Outlines of a theory of the light sense* (L. M. Hurvich & D. Jameson, Trans.): Cambridge, MA: Harvard University Press.
- Irtel, H. (2014). Psychophysical Scaling. In Wiley StatsRef: Statistics Reference Online (eds N. Balakrishnan, T. Colton, B. Everitt, W. Piegorisch, F. Ruggeri and J. L. Teugels). doi:10.1002/9781118445112.stat06504
- Ishihara, S. (2004). *Ishihara's Tests for Colour Deficiency*. Tokyo, Japan: Kanehara Trading Inc.
- Itti, L. (2007). Visual Salience. *Scholarpedia*, 2(9), 3327.
- Jameson, K. A. (2010). Where in the World Color Survey is the support for the Hering Primaries as the basis for Color Categorization? In J. Cohen & M. Matthen (Eds.), *Color Ontology and Color Science* (pp. 179-202). Cambridge, MA: The MIT Press.
- Jameson, K. A., & D'Andrade, R. G. (1997). It's not really red, green, yellow, blue: an inquiry into perceptual color space. In C. N. Hardin & L. Maffi (Eds.), *Color categories in thought and language* (pp. 295-319). Cambridge, UK: Cambridge University Press.
- Judd, D. B. (1951). Report of U. S. Secretariat Committee on Colorimetry and Artificial Daylight (pp. 11). Paris: Bureau Central de la CIE.
- Kay, P., & Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences*, 100(15), 9085-9089.
- Kay, P., & Regier, T. (2007). Color naming universals: the case of Berinmo. *Cognition*, 102(2), 289-298. doi: S0010-0277(05)00244-1 [pii] 10.1016/j.cognition.2005.12.008
- Krauskopf, J., & Gegenfurtner, K. R. (1992). Color discrimination and adaptation. *Vision Research*, 32(11), 2165-2175.
- Kuehni, R. G., Shamey, R., Mathews, M., & Keene, B. (2010). Perceptual prominence of Hering's chromatic primaries. *Journal of the Optical Society of America. A, Optics, image science, and vision*, 27(2), 159-165. doi: 194418 [pii]
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49(2), 467-477.
- Lindsey, D. T., & Brown, A. M. (2006). Universality of color names. *Proceedings of the National Academy of Sciences*, 103(44), 16608-16613.
- Lindsey, D. T., & Brown, A. M. (2009). World Color Survey color naming reveals universal motifs and their within-language diversity. *Proceedings of the National Academy of Sciences USA*. doi: 0910981106 [pii] 10.1073/pnas.0910981106
- Lindsey, D. T., Brown, A. M., Brainard, D. H., & Apicella, C. L. (2015). Hunter-Gatherer Color Naming Provides New Insight into the Evolution of Color Terms. *Current Biology*, 25(18), 2441-2446. doi: 10.1016/j.cub.2015.08.006
- Lindsey, D. T., Brown, A. M., Brainard, D. H., & Apicella, C. L. (2016). Hadza Color Terms Are Sparse, Diverse, and Distributed, and Presage the Universal Color Categories Found in Other World Languages. *i-Perception*, 7(6), 2041669516681807. doi: 10.1177/2041669516681807
- Lindsey, D. T., & Brown, A. M. (2019). Recent progress in understanding the origins of color universals in language. *Current Opinion in Behavioral Sciences*, 30, 122-129. doi:10.1016/j.cobeha.2019.05.007
- Logvinenko, A. D. (2012). A theory of unique hues and colour categories in the human colour vision. *Color Research & Application*, n/a.
- Logvinenko, A. D., & Beattie, L. L. (2011). Partial hue-matching. *Journal of Vision*, 11(8), 1-16.
- Mizokami, Y., Werner, J. S., Crognale, M. A., & Webster, M. A. (2006). Nonlinearities in color coding: compensating color appearance for the eye's spectral sensitivity. *Journal of Vision*, 6(9), 996-1007. doi: 10.1167/6.9.12
- Munsell, A. H. (1912). A Pigment Color System and Notation. *The American Journal of Psychology*, 23(2), 236-244. doi: 10.2307/1412843
- O'Neil, S. F., McDermott, K. C., Mizokami, Y., Werner, J. S., Crognale, M. A., & Webster, M. A. (2012). Tests

- of a functional account of the Abney effect. *J Opt Soc Am A Opt Image Sci Vis*, 29(2), A165-173. doi: 226815 [pii]
- Olkkonen, M., Witzel, C., Hansen, T., & Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. *Journal of Vision*, 10(9), 1-22. doi: 10.9.16 [pii] 10.1167/10.9.16
- Pastilha, R. C., Lihares, J. M. M., Rodrigues, A. I. C., & Nascimento, S. M. C. Describing natural colors with Munsell and NCS color systems. *Color Research & Application*, 0(0). doi: doi:10.1002/col.22355
- Philipona, D. L., & O'Regan, J. K. (2006). Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neuroscience*, 23(3-4), 331-339. doi: S0952523806233182 [pii] 10.1017/S0952523806233182
- Regier, T., Kay, P., & Cook, R. S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences*, 102(23), 8386-8391. doi: 0503281102 [pii] 10.1073/pnas.0503281102
- Regier, T., Kay, P., & Khetarpal, N. (2007). Color naming reflects optimal partitions of color space. *Proceedings of the National Academy of Sciences USA*, 104(4), 1436-1441. doi: 0610341104 [pii] 10.1073/pnas.0610341104
- Rosch Heider, E. (1972). Universals in color naming and memory. *Journal of Experimental Psychology*, 93(1), 10-20.
- Siuda-Krzywicka, K., Boros, M., Bartolomeo, P., & Witzel, C. (2019). The biological bases of colour categorisation: From goldfish to the human brain. *Cortex*. doi:https://doi.org/10.1016/j.cortex.2019.04.010
- Sturges, J., & Whitfield, T. W. A. (1997). Salient features of Munsell colour space as a function of monolexic naming and response latencies. *Vision Research*, 37(3), 307-313. doi: S0042-6989(96)00170-8 [pii]
- Valberg, A. (2001). Unique hues: an old problem for a new generation. *Vision Research*, 41(13), 1645-1657. doi: S0042-6989(01)00041-4 [pii]
- Vazquez-Corral, J., O'Regan, J. K., Vanrell, M., & Finlayson, G. D. (2012). A new spectrally sharpened sensor basis to predict color naming, unique hues, and hue cancellation. *Journal of Vision*, 12(6), 7. doi: 10.1167/12.6.7
- Vos, J. J. (1978). Colorimetric and photometric properties of a 2° fundamental observer. *Color Research & Application*, 3(3), 125-128. doi: 10.1002/col.5080030309
- Webster, M. A., Webster, S. M., Bharadwaj, S., Verma, R., Jaikumar, J., Madan, G., & Vaithilingham, E. (2002). Variations in normal color vision. III. Unique hues in Indian and United States observers. *Journal of the Optical Society of America A*, 19(10), 1951-1962.
- Weiss, D., Witzel, C., & Gegenfurtner, K. (2017). Determinants of Colour Constancy and the Blue Bias. *i-Perception*, 8(6), 2041669517739635. doi: 10.1177/2041669517739635
- Witkowski, S. R., & Brown, C. H. (1982). Whorf and Universals of Color Nomenclature. *Journal of Anthropological Research*, 38(4), 411-420.
- Witzel, C. (2016). New Insights Into the Evolution of Color Terms or an Effect of Saturation? *i-Perception*, 7(5), 1-4. doi: 10.1177/2041669516662040
- Witzel, C. (2018a). Misconceptions About Colour Categories. *Review of Philosophy and Psychology*, 1-42. doi: 10.1007/s13164-018-0404-5
- Witzel, C. (2018b). The role of saturation in colour naming and colour appearance. In L. W. MacDonald, C. P. Biggam & G. V. Paramei (Eds.), *Progress in Colour Studies: Cognition, language and beyond* (pp. 41-58). Amsterdam/Philadelphia: John Benjamin Publishing Company.
- Witzel, C. (2019). Variation of saturation across hue affects unique and typical hue choices. *i-Perception*, 10(5), 1-14. doi: 10.1177/2041669519872226
- Witzel, C., Cinotti, F., & O'Regan, J. K. (2015). What determines the relationship between color naming, unique hues, and sensory singularities: Illuminations, surfaces, or photoreceptors? *Journal of Vision*, 15(8), 19. doi: 10.1167/15.8.19
- Witzel, C., & Franklin, A. (2014). Do focal colors look particularly "colorful"? *Journal of the Optical Society of America. A, Optics, image science, and vision*, 31(4), A365-374. doi: 10.1364/JOSAA.31.00A365
- Witzel, C., & Gegenfurtner, K. R. (2013). Categorical sensitivity to color differences. *Journal of Vision*, 13(7). doi: 10.1167/13.7.1
- Witzel, C., & Gegenfurtner, K. R. (2015). Chromatic Contrast Sensitivity. In R. Luo (Ed.), *Encyclopedia of Color Science and Technology* (pp. 1-7). Berlin: Springer.
- Witzel, C., & Gegenfurtner, K. R. (2016). Categorical perception for red and brown. *Journal of Experimental Psychology: Human Perception & Performance*, 42(4), 540-570. doi:10.1037/xhp0000154
- Witzel, C., & Gegenfurtner, K. R. (2018a). Are red, yellow, green, and blue perceptual categories? *Vision*

- Research, in press. doi:  
<https://doi.org/10.1016/j.visres.2018.04.002>
- Witzel, C., & Gegenfurtner, K. R. (2018b). Color Perception: Objects, Constancy, and Categories. *Annual Review of Vision Science*, 4(1), 475-499. doi: 10.1146/annurev-vision-091517-034231
- Witzel, C., Maule, J., & Franklin, A. (2013). Focal colors as perceptual anchors of color categories. *Journal of Vision*, 13(9 VSS abstracts), 1164-1164. doi: 10.1167/13.9.1164
- Witzel, C., van Alphen, C., Godau, C., & O'Regan, J. K. (2016). Uncertainty of sensory signal explains variation of color constancy. *Journal of Vision*, 16(15), 8. doi: 10.1167/16.15.8
- Wool, L. E., Kombar, S. J., Kremkow, J., Jansen, M., Li, X., Alonso, J. M., & Zaidi, Q. (2015). Salience of unique hues and implications for color theory. *Journal of Vision*, 15(2). doi: 10.1167/15.2.10
- Yendrikhovskij, S. N. (2001). Computing color categories from statistics of natural images. *Journal of Imaging Science and Technology*, 45(5), 409-417.
- Zaslavsky, N., Kemp, C., Regier, T., & Tishby, N. (2018). Efficient compression in color naming and its evolution. *Proceedings of the National Academy of Sciences*. doi:10.1073/pnas.1800521115